Riskalyzer: Inferring Individual Risk-Taking Propensity Using Phone Metadata

VIVEK K. SINGH, Rutgers University and Massachusetts Institute of Technology, USA
RUSHIL GOYAL, Rutgers University, USA
SHENGHAO WU, Columbia University, USA

An individual's risk-taking propensity is "the stable tendency to choose options with a lower probability of success, but greater rewards". This risk propensity plays a central role in decision making by customers as well as managers, and is a mediator in behavior associated with security, privacy, health, finance, and well-being. Most common approach to understanding an individual's risk propensity remain lab-based games and surveys. Administering such surveys and games is a manual, time, and money intensive process that is also fraught with multiple biases. Recently, smartphones are increasingly seen as large-scale sensors of human activity, recording data related to physical and social aspects of people's lives. Building on this trend we investigate the potential of passive phone-based data for automatically inferring an individual's risk propensity. Specifically, we describe a novel approach to model an individual's risk propensity based on her mobile phone usage. Based on a 10-week field + lab study involving 50 participants, we report that: (1) multiple phone-based features (e.g., average gyradius) are intricately associated with participants' risk propensity; and (2) a phone-based model outperforms demography-based models by 39% in terms of accuracy of predicting risk propensity. In organizational terms, a better understanding of risk behavior could contribute significantly to risk management programs. At the same time, such results could open doors for more nuanced understanding of the underlying human risk phenomena and their interconnections with social and mobility behavior.

CCS Concepts: • **Human-centered computing → Ubiquitous computing**; **Mobile computing**;

Additional Key Words and Phrases: Risk; Risk propensity; Mobile Sensing; Behavioral Informatics; Phoneotype; Phone metadata

ACM Reference Format:
Vivek K. Singh, Rushil Goyal, and Shenghao Wu. 2018. Riskalyzer: Inferring Individual Risk-Taking Propensity Using Phone Metadata. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 2, 1, Article 34 (March 2018), 21 pages. https://doi.org/10.1145/3191766

1 INTRODUCTION

An individual's tendency to behave riskily, or to opt for a bigger reward with higher chance of failure sometimes enables her obtain better results [33]. Risk taking behavior has thus been actively studied over past decades (e.g., [33, 53]). Such literature examines risk propensity through either empirical or theoretical ways and provide insightful points on the assessments of it. Understanding risk propensities of individuals is central to multiple sub-fields in behavioral economics, e-commerce, health care, security, and management information systems

Authors' addresses: Vivek K. Singh, Rutgers University and Massachusetts Institute of Technology, 4 Huntington Street, New Brunswick, NJ, 08901, USA; Rushil Goyal, Rutgers University, 4 Huntington Street, New Brunswick, NJ, 08901, USA; Shenghao Wu, Columbia University, Math & Statistics, 601 W. 112th St. New York, NY, 10025, USA, sw3002@columbia.edu.

2474-9567/2018/3-ART34 $15.00
https://doi.org/10.1145/3191766

[28, 35, 40, 47, 61, 66, 69, 87, 91]. In fact, risk analysis is an evolving and growing field in its own right. For example, risk propensity is an important determinant of decision making behavior by managers as well as customers [36, 43, 78]. Similarly, risk has been identified as an important mediator for individual information seeking behavior (e.g., in health or financial settings) [64, 67]. Further, risk propensity is an important component of modeling the user in studies related to privacy, trust, and security [22, 37, 65].

Most of the assessment of risk taking behavior lies inside the framework of experiment and manual observation. A small number of participants are usually gathered to participate in some games designed specifically to observe their tendency to behave riskily [6, 53]. Other methods connect risk tendency demographic and domain descriptors (e.g. gender, immigrant identity) [34, 85]. Unfortunately, the human-related information captured by observations in limited, unnatural settings must contend with multiple challenges, including subjectivity in observation, biases, and limited observation opportunities and at the same time involve substantive manual effort, cost and time [17, 26]. Further, the reliance only on laboratory elicitable characteristics to identify risk propensities hampers the growth of the field of risk analysis. Such an approach essentially rules out the potential for identifying behavioral markers based on communication or mobility traces spread over space and time (e.g., diversity of locations visited) to predict individual propensities to take risk.

No existing research has used phone metadata (e.g., number of phone calls made per week, location diversity) to study an individual risk propensity. This is surprising, given the frequent associations reported between social behavior and personality with risk taking [53, 86], and recent efforts that have connected, both social phenomena (e.g. availability of social capital) and personality traits with phone metadata [11, 16, 39].

We explore the associations between an individual's risk propensity (as assessed through a "gold-standard" survey) and several phone-based features and then build a machine learning model to accurately predict an individual's risk propensity score. This builds upon a recent line of work in ubiquitous computing literature that utilizes *phoneotypes* i.e. composites of an individual's traits as observable via a mobile phone, to automatically infer an individual's characteristics (e.g. cooperation) [75].

This paper seeks to make two contributions:

1. To motivate and ground the use of phone-based signals for inferring individual risk propensities; and
2. To define a machine learning model to automatically infer an individual's risk propensity.

Several commercial, social, and citizen organizations could benefit from the proposed approach. For example, a health informatics professional working on smoking cessation could create different interventions for different individuals based on their risk-taking propensity. Similarly, risk researchers could also take advantage of this approach to obtain larger, longer-term, data corpus by asking participants to install an application into their smartphone, which could help identify long-term behavioral predictors of risk propensity (e.g. diversity of locations visited), which could simply not be investigated using occasional in-lab studies. Lastly, individuals could receive customized suggestions in multiple settings (finance, investments, security) based on their automatically-inferred risk propensities.

The organization of the rest of the paper is as follows: We first provide the relevant literature on the definition, characteristics, and measurement of risk propensity which provides the theoretical groundwork for our research. Then the description of a 10-week wellbeing study with 50 participants is provided. Next, we discuss the features used in this study and explain the rationale behind them. This is followed by the results obtained through regression, classification, and a post-hoc correlation analysis. The implications of the result are discussed along with the limitations and potential for future work.

2 BACKGROUND AND RELATED WORK

Risk propensity has been a core issue investigated in multiple social and behavioral sciences, but its definition and measurement remains debatable [53]. In this section we summarize the related work under four dimensions.

First we summarize the literature on risk behavior as a field of study with focus on the conceptualization of risk and risk taking behavior. This provides the theoretical background for our analysis. Next we summarize the existing methods of assessing and measuring risk (e.g., the personality measurement and the behavior-based measurement), which motivates our novel approach of utilizing phone metadata and provides a baseline for comparison. Next, we summarize the literature on social capital and discuss risk in the context of social capital and social networks, as much of the phone metadata serves as an indicator of individuals' social capital and can hence influence their risk propensity. We use social capital as the key theoretical construct that connects phone-behavior and risk propensity. Lastly, we discuss recent literature on the applications of mobile phone-based data to model and predict different individual traits using statistical learning methods, which provides the empirical foundation of our analysis on the phone-based data.

2.1 Risk as a Field of Study

Given the wide variety of subject areas that study risk, there is also a variation in the definitions adopted. For example, Sim and Amy (1992) suggest that three facets of risk are key to the understanding of its concept: outcome uncertainty, outcome expectations, and outcome potential [77]. They further associate risky behavior with the extent of risk of a decision. Connecting to the three dimensions of risk, they define a risky behavior as "(a) their expected outcomes are more uncertain, (b) decision goals are more difficult to achieve, or (c) the potential outcome set includes some extreme consequences" [77]. Bechara [6] described risk-taking as a "voluntary pursuit of activities" that causes losses. Moore and Gallone define risk as the trade-off between short-term gain and long-term loss [51].

There exist different schools of thoughts on the processes surrounding risk and risky behavior. Different explanations of its mechanics (e.g., as a trait vs. a state) exist [79, 88], and the emerging literature suggests that neither of these can provide the complete picture, but rather risky behavior is a complex composite of both the individual's stable traits and as well as the state they find themselves in [79, 80]. While acknowledging the role played by the situational context, the focus of this work remains on identifying the trait like mechanisms (i.e. stable risk propensities), which presumably manifest themselves in long-term patterns of human behavior as observed via mobile phones. Hence, in this paper we focus on studying risk propensity, defined as "the stable tendency to choose options with a lower probability of success, but greater rewards" [2].

2.2 Measures of Risk

Measuring risk propensity has attracted the attention of researchers in various disciplines including behavioral economists (e.g. [69]); consumer research (e.g. [28]); cognitive psychologists (e.g.[35, 40, 47]); social psychologists (e.g. [87, 91]); as well as financial analysts and financial planners (e.g. [61, 66]). However, the study and assessment of risk tolerance is fragmented, with each discipline using different methodologies and with different focuses ([29]).

While some researchers look at factors such as demographics, socio-economic status, personality, attitudes about money and even birth order to predict individual financial risk tolerance (e.g. [28, 29, 82, 87]); others look at how framing, contextual factors and situational factors influence risk tolerance (e.g. [45, 69]). The assessment of risk propensity (and behavior) has also proven to be difficult due to the subjective nature of risk taking, and has been open to the interpretations of researchers within their fields of work. Behavioral economists typically measure risk tolerance by the choices the individual make in financial situations, for example, Roszkowski and Snelbecker (1990) looked at how framing affects "risk tolerance" as measured by the choice of either certain gain or probable gain. Other efforts have used tasks that mimic gambling or those which create analogs to smoking [6, 44, 79] for prediction of risk-taking. While great at eliciting user behavior, such lab-based tasks are often critiqued to be costly and too closely tied to a single application domain.

Other researchers have measured risk by designing questionnaires containing behavioral statements and personality-style questions [27, 70, 90]. While, such survey instruments are easier to administer and can potentially capture more abstract representations of individual propensities, there is often a gap reported in user intentions and behavior. Personality oriented risk modeling has been very actively studied in multiple social and biological sciences. For example, the five-factor personality model has shown power in predicting risk propensity [51, 53]. In fact, other type of markers including measuring individual testosterone levels [4] and FMRI indicators [63] have also been used to identify risk propensity.

Given, the wide variety of options and a lack of consensus, for this exploratory work we focus on a simple to administer survey-based measure to characterize risk propensity (details follow). The key goal of this work is to study the interconnections between phone based data and individual risk propensity, and the results could motivate future studies that characterize risk propensity and behavior in other ways.

2.3 Risk and Social Capital

An individual's risk taking behavior is often related to their social behavior [7, 49, 68]. A very important concept in the study of social behavior is that of social capital [59, 60]. Putnam [60] characterizes social capital as trust, network structures, and norms that promote cooperation among actors within a society for their mutual benefit. He also suggests that formal membership, civic participation, social trust, and altruism are indicators of social capital [59]. A strong line of research has showed the predictive power of dimensions of an individual's social capital on her risk propensity and risky behavior [7, 15, 49, 68].

However, there is contradicting evidence on the effects of social capital on risk-taking propensity as well as outcomes. On one hand, multiple studies have connected higher social capital with higher risk-taking propensity. For example, a study on Chinese immigrants entrepreneurs has found that those with greater social capital have a higher risk-taking propensity in their business activity [68]. Another research argues that social capital plays a central role in the ability of young people to navigate risk decisions [7] and [15] suggests that social capital plays an important role in predicting the success of nascent business ventures. On the other hand, higher social capital has also been associated with lesser odds of engaging in risky behavior. While, [49] identifies a negative association between social capital and violence and alcohol assumption by young people, a study by Crosby et al. [14] suggests that social capital is inversely correlated with sexual risk behaviors.

Hence, while there is enough evidence to suggest that there is some association between social capital and risk propensity, nuanced work, and scalable analysis methods might be useful in creating a more comprehensive picture. Therefore, looking beyond aggregate social capital score, in this work we would also consider the *type* of social capital (bridging vs. bonding) available to the individuals in studying the associations between risk and social capital.

Recent studies have connected social capital with phone use behavior [11, 39, 58], thus suggesting that phone use behavior could also be predictive of an individual's risk propensity.

2.4 Use of Mobile Phone to Predict Health and Wellbeing

The advancement of data collection (using phone sensors) and data analysis tools has made it possible to study individual's behavior outside of the laboratories and in the long-term. Rich schools of research have succeeded in predicting demographics, personality and cooperation propensities using phone-generated metadata[16][21] [75]. Such data have also been connected with different measures of health and well-being in recent ubiquitous computing literature [3, 62, 84]. Lastly, prior research has also shown the predictive power of mobile phone data on individuals' social capital [11, 39, 58], which we consider to be a key connector between phone-based behavior and risk propensity. These efforts motivate the use of phone data to predict risk propensity.

3 STUDY

We probe the interconnections between cooperation levels and phone-based metrics based on a ten-week field and lab study.

3.1 Participants

The participants for the study were recruited using flyers, email announcements, and social media posts in the area surrounding a major North American university. A total of 59 participants completed the study. Some of the participants did not complete all the surveys, and some did not enter their unique identifying code consistently across surveys resulting in a dataset of 54 participants (Off-by-one errors in the 14/15 digit IMEI codes were corrected manually.) Of these 54 participants two had multiple missing values and two had abnormally low location counts (presumably due to disabling location services on their phone) and hence were removed from the dataset. This resulted in a dataset of 50 participants used in the rest of this paper.

Out of the 50 participants in the resulting dataset, 33 were male, and 17 were female. The most common age group was 18-21 years, most common marital status was single, the most common education level was "some college," and the median annual family income was in the range US $35,000- $49,999.

3.2 Method

The data used in this paper were obtained as part of the Rutgers Well-being study. In this study, the participants were invited to attend three in-person sessions which involved filling out several surveys pertaining to health, well-being, cooperation and some demographic data. The three surveys of relevant to the current paper are those on risk-propensity, demography, and personality traits.

The participants had to be above 18 years of age, conversant in English, Android phone users on a daily basis, and willing to travel to Rutgers (New Brunswick, NJ) campus for three in-lab sessions. In the initial session, the participants were asked to read and sign a consent form to join the study and install a study client app onto their smartphones. The study client app was built using Funf platform [1] and a screenshot is shown in Figure 1. The app was tested to be compatible with Android OS versions 3.0 though 5.0. Android version 5.0 was the latest version at the time of conducting this study (Spring 2015).

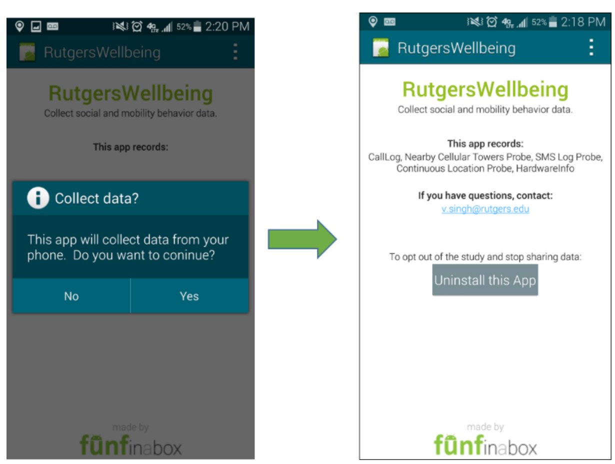

Fig. 1. Screenshot of the study client app

The study started with an orientation program where participants were trained on how to use the app. Participants could opt out of the study at any point. The survey order was randomized for different participants. The participants were also instructed on the process of uninstalling the app and stopping data collection in the exit session.

All personnel involved in the study underwent human subject training and Institutional Review Board certification. In order to protect participants' personal information all participants were identified using their phone's unique IMEI number which was kept constant throughout the course of the study. This unique identifier was anonymized via one-way cryptographic hashing before further data analysis was undertaken.

3.3 Data Collection

Following information were collected from the participants for the ten week period:

- Communication log data: This included date and time of communication event, the type of communication (text/call), incoming or outgoing, duration, and anonymized identifiers for both sender and receiver. Actual call audio and message text were **not** collected in this study.
- Physical location: This included the location of the device using GPS and approximate location area based on the cell tower. Such data were collected at an hourly resolution to a trade-off between the battery consumption and the detail of coverage. This work focuses on GPS data that was analyzed as a (latitude, longitude) tuple at the fourth decimal point resolution.[1]

3.4 Survey Measures

We consider three different survey measures in this work. The main target variable of this work is individual risk-taking propensity, a quantification of which will be compared with various phone based features. We also compare the performance of phone-based models with two baseline (and potentially complementary) types of information - personality and demographic data.

Risk Propensity Survey: We considered multiple survey options for measuring risk propensity [27, 56, 70, 90]. While multiple options were available, we focused on a metric that would be easy to execute, widely utilized in the literature, and applicable to a relatively broad range of every-day scenarios. Based on these considerations, we decided to utilize the risk propensity scale defined by Pan and Zinkhan [56]. This survey was easy to execute (consists of 5 questions) and has been widely used by other studies in the literature [22, 50]. The questions in the survey included 'To achieve something in life, one has to take risks' and 'If there was a great chance to multiply my earnings, I would invest my money even in the shares of a completely new and uncertain firm.' Thus, while they were leaning towards financial decisions, they were still general enough to be interpreted in terms of everyday life decisions. As mentioned earlier, the key goal of this work is to study the interconnections between phone based data and risk propensity, and the results could motivate future studies that characterize risk propensity and behavior in other ways.

A risk score is calculated based on the intent of agreement to different statements in the survey. For each of the statements the participants could choose from a 5-point Likert scale representing Strongly Disagree (1) to Strongly Agree (5). A higher score indicated a higher propensity to engage in risky behavior.

Demographic Descriptors: The participants were also asked questions about their demography (age, gender, marital status, level of education, and family income level) using another survey.

[1]This roughly corresponds to 10m by 10m blocks and was selected as the scale since current GPS data have a pseudo-range accuracy of about 8 meters at 95% confidence level(http://www.gps.gov/systems/gps/performance/accuracy). Since many of the participants spent inordinate amounts of times at the same location, while we were interested in characterizing their movement patterns, many of the features focused on the unique locations visited by the users.

Personality Descriptors: Prior literature has reported that the five-factor personality model has predictive power on risk propensity [51, 53, 54]. Hence, we obtained the big-five personality scores for the participants using the well-established survey defined by John and Srivastava [38].

3.5 Smartphone Data Measures

Based on a survey of existing literature we have identified a set of features to provide a representation of a user's social- mobile behavior.

These features are based on existing literature with the inspirations coming from both conceptual and empirical literature. From a conceptual perspective, these features are inspired by the literature on social capital [60] and its variants on 'bridging' and 'bonding‘' capital [86]. As suggested by Williams[31], we connect the concepts of 'bonding' and 'bridging' to those of 'strong' and 'weak' ties as proposed by Granovetter and other researchers[31][55][52]. All these features are based on a key working assumption that individual traits manifest themselves in the long-term socio-mobile behavior patterns of the users [72, 75], and might be predictive of one's self-reported propensity to take risks. From an empirical perspective, these features are based on multiple recent studies that have also tried to quantify long-term socio-mobile behavior of individuals with a goal to identify per-sonal traits and well-being [3, 16, 62, 71, 75, 84]. In defining these features (total 23), we integrate some of the more obvious usage patterns (total number of calls/sms messages) with more nuanced/exploratory aspects of phone usage while still keeping the number of features manageable.

A summary of the features considered in this study is presented in Table 1.

3.5.1 Activity Level (Call, SMS, Location). The overall social activity is measured by the number of calls, SMS, and unique locations experienced by a person over the course of the study. High count would imply that the person is more actively involved with people and places around her, and thus more socially engaged [13, 19]. Since social activity level is related to the degree of the individual's civic participation level, this feature could be used to indicate her social capital [59] and thus her risk propensity[68].

$$SA_i = \sum \text{Activity Levels}$$

We observe the physical location of each user once per hour. Hence counting the number of locations per day would yield exactly the same value (24) for every user. Hence, we focus on unique locations visited by the users.

3.5.2 Reciprocity. Along with the frequency of communication, the ease with which communication is conducted can also be a significant predictor of an individual's social capital [32, 60]. People who are easy to contact can be said to engage with a wider social network than those who remain unavailable. Here, ease of access is operationalized using features related to accepting or rejecting calls [16, 24]. This includes the ratio of incoming to outgoing calls, percentage of calls missed, and call response rate (percentage of missed calls responded back). These features taken together provide an understanding of how keen is the individual to engage with calls initiated by others in the network.

In Out Ratio (Call, SMS).

$$IOR = \frac{\text{Incoming communication count}}{\text{Outgoing communication count}} * 100$$

It is defined as the ratio of the number of incoming Calls/SMS to outgoing Call/SMS messages. A higher In-Out ratio would suggest that the individual is more likely to receive messages and/or reply back to the messages received. This could also indicate the individual's civic participation level and suggest her social capital volume[59].

Missed Call Percentage.

$$M_i = \frac{\text{Missed incoming call count}}{\text{Total incoming call count}} * 100$$

Missed call percentage quantifies how reluctant the user is to respond to the call. Higher missed call percentage could indicate weaker social capital (less civic participation level [59]) and higher prudence which in turn is connected to less trust and social capital[23].

Response Rate (Call).

$$R_i = \frac{\text{Responded incoming call count}}{\text{Total incoming call count}} * 100$$

We define a communication (call, SMS) as responded when the person calls back a missed call within one hour of its receipt.

Even though this number might be impacted by the frequency the person checks the phone, it indicates the person's tendency to respond to uncertainty (the unknown content of the call and the possibly unknown identity of the caller). Such a tendency to maintain social connections is associated with social capital, which in turn may be associated with risk propensity [16, 32].

3.5.3 Tie Strength. Significant prior literature connects strong and weak ties with social capital [32, 46]. For both calls and SMS, we quantify the relative percentage of interactions that take place with these "strong ties" as well as "weak ties".

Strong Ties Engagement Ratio (Call, SMS).

$$S_i = \frac{\text{Communication count for the top third contacts}}{\text{Total communication}} * 100$$

This quantifies the percentage of communication effort that a user devotes to her preferred contacts. A high strong tie score would imply that the user engages much more with her strong ties compared to other contacts. A user who stays with her strong ties tend to be more conservative and careful in making a decision, thus has less risk propensity. Strong ties can also be represented by one's social capital [60] and is also implicitly related to one's risk propensity.

Weak Ties Engagement Ratio (Call, SMS).

$$W_i = \frac{\text{Communication count for the bottom third contacts}}{\text{Total communication}} * 100$$

This quantifies the percentage of communication effort that a user devotes to her less preferred (bottom third) contacts. A high weak tie score would imply that the user spends significant communication effort to interact with even the less preferred of his contacts.

Strong ties and weak ties together may affect, in positive and negative way[18], the user's entrepreneurship which is connected the person's personality and social capital[20].

3.5.4 Exploration and Novelty. The relationship between a diverse network and an increased social capital is well documented in literature [32]. Prior literature has also connected phone-based diversity with an individual's propensity to cooperate [75]. Further, an important aspect of social capital is the growth of networks. Hence, we also consider "new contacts" i.e. those who are not present in the first four weeks of the data collection period. We characterize the number of new contacts, number of such new contacts for outgoing calls, and the time spent in talking to these new contacts.

Entropy (Call, SMS, GPS).

$$E_i = \sum_j p_{ij} log_b p_{ij}$$

Where p_{ij} is the percentage of calls made by individual i to contact j, and 'b' is the total number of contacts. This is also called the Shanon Entropy [73] which quantifies the uncertainty of the information. A user with higher entropy is more likely to distribute her time among every contact and gives no priority to anyone.

Time Spent with New Contacts (Call).

$$T_i = \sum_j t_{ij}$$

Where t_j is the time (in seconds) spent by user i while talking to her j_{th} new contact. A contact is considered "new" if the contact does not appear in the participant's logs for the first four weeks. A higher value indicates the user's more investment on her new contacts compared to other users, and thus her more trust and social engagement which is connected to high social capital[59][23].

Percentage of Time Spent with New Contacts (Call).

$$PN_i = \frac{\text{Time spent with new contact}}{\text{Total communication time}}$$

Similar to the previous feature, this feature measures in ratio scale how much time the user devotes to her new contacts as compared to her frequent contacts. This could be more indicative of the users' risk propensity as it is on a [0,1] scale which makes it more comparable than the absolute value.

Gyradius Average(Location).

$$G_i = \frac{\sum_j s_{ij}}{n_i}$$

Where s_{ij} is the user j's current distance to her home. n_j is the totally number of scans on the user's GPS location. A high average gyradius indicates a large activity radius of the user, which could be a useful indicator of the user's exploration intensity.

Returner1 vs Explorer0 (Location).

$$\text{Characteristic Distance} = \sqrt{\frac{1}{N} \sum_{i \in \{1,2\}} n_i (r_i - r_{cm})^2}$$

where $N = n_1 + n_2$
n_i is the frequency of the i-th frequently visited place
r_i is the 2-d location of the i-th frequently visited place
r_{cm} is the 2-d location of the centroid of all the user's trajectories.
Inspired by [57], we label a user as explorer (with value close to 0) when she tends to visit a large number of different places and returner (with larger value) when she performs recurrent visitation to a small number of destinations. The returner(1) vs. explorer(0) value is determined by her characteristic distance which is the standard deviation of the user's GPS location of the first and second frequently visited locations with her geography centroid as the mean.

3.5.5 Temporal Rhythms.

Table 1. A summary of the features considered in this study

Type of feature	Literature Support	Feature	Data Type	Mean	Median
Activity Levels	Conceptual: [19, 59, 68]	Call_Count	Call	513.7	314.0
	Empirical: [13, 16, 24, 84]	Sms_Count	SMS	3611.6	2553.5
		Unique_location_count	GPS	277.1	288.0
Reciprocity	Conceptual: [32, 46, 68]	InOut_ratio_Call	Call	63.3	53.3
	Empirical: [16, 25, 74]	Inout_ratio_SMS	SMS	122.5	113.0
		Missed_call_percentage	Call	16.9	16.0
		Call_Response_rate	Call	37.2	37.4
Tie Strength	Conceptual: [32, 46, 60]	Call_Strong_Ties_Engagement_Ratio	Call	82.3	84.0
	Empirical: [25, 74, 75]	Sms_Strong_Ties_Engagement_Ratio	SMS	88.6	90.5
		Call_Weak_Ties_Engagement_Ratio	Call	5.9	5.0
		Sms_Weak_Ties_Engagement_Ratio	SMS	2.1	1.0
Exploration and Novelty	Conceptual: [32, 41, 60]	Call_Entropy	Call	65.7	59.0
	Empirical: [16, 17, 57, 75]	Sms_Entropy	SMS	63.0	67.5
		Location_Entropy	GPS	75.7	76.0
		Time_spent_with_new_contacts (seconds)	Call	7413.9	3785.5
		Percentage_of_time_new_contacts	Call	15.0	11.3
		Gyradius_Average	GPS	101.7	11.9
		Returner1_vs_Explorer0	GPS	0.24	0.02
Temporal Rhythms	Conceptual: [5, 10, 42]	DN12am_Call	Call	4.3	3.6
	Empirical: [3, 16, 71, 76]	DN12am_Sms	SMS	3.9	3.2
		Weekday_weekend_call_ratio	Call	2.5	2.3
		Weekday_weekend_SMS_ratio	SMS	4.0	3.2
		Regularity_AR31	Call	0.001	0.01

Diurnal Activity Ratio (Call, SMS).

$$\text{Diurnal Ratio} = \frac{\text{Communication count during phase1}}{\text{Communication count during phase2}}$$

This metric measures the circadian rhythms as the ratio of activity between two 12 hours "phases" commonly called AM and PM i.e. AM = 00:00 am to 11:59 am and PM = 12:00 pm to 11:59 pm. Here, the more active phase i.e. PM is considered as phase 1. The individuals who have a diurnal activity ratio closer to one do not prioritize communication or travel during specific hours of the day and spread out their social interactions unbiased of their personal circadian rhythms, while others give higher priority to their personal rhythms or have personal or social preferences resulting in different activity levels during different phases. In prior work, such a ratio has been found to be closely related to the individual's impulsiveness[42] [5] and to also have a relationship to an individual's risk propensity i.e. a morning person is more risk-averse [42].

Weekday-weekend Ratio (Call, SMS).

$$W_i = \frac{\text{Communication count on weekdays}}{\text{Communication count on weekends}}$$

A high weekday-weekend ratio indicates that the user has less communications during the weekends. The high ratio could be due to the user's inactivity during her spare time. There is no specific literature that ties weekend vs. weekday communication ratio to risk propensity, but this feature has been used in multiple prior studies quantifying phone use behavior (e.g., [16]).

Regularity (Autoregressive coefficient 31 days) (Call). We model the regularity for an user individual based on ARIMA (autoregressive integrated moving average) time-series coefficients. This model takes the form:

$$X^t = c + \sum_{i=1}^{p} \phi_i \cdot x^{t-1} + \epsilon^t$$

where X^t is the number of phone calls made during time period t, c is a constant and ϵ^t is the noise. Specifically, we define regularity scores based on ϕ_i coefficients quantifying the degree to which knowing a person's number of calls made on an "earlier" day (here, previous month) can predict their number of phone calls for today. In effect, this represents the repetition in the user's behavior in time period of a month.

The 50 participants considered in this study made a total of 25,683 calls, exchanged 180,579 messages, and visited 13,855 unique locations during the 10 week period. This corresponds to an average of 51 (median = 31) calls per user per week, 361 (median = 255) text messages, and 28 (median = 29) unique locations visited per user per week. A summary of all the features considered in this study, and their observed values (per user over the ten week period) is shown in Table 1.

4 RESULTS

While certain applications may require prediction of exact numeric risk propensity score, others might need to work with broader classifications e.g. "high' or "low" risk propensity. Hence in order to support both these kind of applications, and to also obtain higher confidence in the observed results, we decided to undertake both these kinds of analysis.

4.1 Regression

We undertook OLS (Ordinary Least Squared) regression on the different subsets of the data to test what percentage of variance in the response variable can be explained by the considered covariate data. Specifically, we tried 6 different data models for comparison: the social-mobile features only, the demographic features (age, gender, grade, self-description, family income) only, the personality features (extraversion, agreeableness, conscientiousness, neuroticism and openness) only, social-mobile + demographic, social-mobile+personality and social-mobile + demographic + personality. All the continuous variables were standardized to mean 0 and standard deviation 1 and all the categorical variables were entered as dummy variables. For each aforementioned model, the features were selected using backward stepwise elimination [89].

The two non-phoneotype feature based approaches (Only Demography and Only Personality) served as two baselines for us to compare and evaluate the explanatory power of the social-mobile features. All the calculations were executed using R (version 3.2.3).The amount of variance explained by the corresponding model(adjusted R^2), the selected features and the corresponding significance (p value) for the model are shown in Table2.

From Table 2 it can be seen that demographic and personality features alone were not significantly predictive of a person's risk propensity. They yielded adjusted R-squared of only 0.042 and 0.049 respectively. On the other hand, The phoneotype features (Call, SMS, GPS) have a moderate power in predicting one's risk propensity: an adjusted R squared of 0.499 shows that around 49.9% of the variance in the response variable can be explained by the phone features. This is a reasonably strong result, especially in comparison with the results with using demographic features only.

Table 2. Different regression models considered for explaining risk propensity including the features utilized and the variance explained.

Model	Selected features	Adjusted R-squared	Significance(p-val)
Only Phoneotype	Call Count, Sms_Count, InOut_ratio_Call, Time_spent_with_new_contacts, Percentage_of_time_spent_with_new_contacts, Location_Entropy, Call_Strong_ties, Sms_Strong_ties, Sms_Weak_ties, DN12am_Sms,Gyradius_Average, Returner1_vs_Explorer0	0.499***	0.00006
Only Demographic	Grade, Race	0.042	n.s.
Only Personality	Agreeableness, Openness	0.049	n.s.
Phoneotype+Demographic	Call Count, Sms_Count, InOut_ratio_Call, Time_spent_with_new_contacts, Percentage_of_time_spent_with_new_contacts, Location_Entropy, Call_Strong_ties, Sms_Strong_ties, Sms_Weak_ties, DN12am_Sms, Gyradius_Average, Returner1_vs_Explorer0, Age	0.505***	0.00008
Phoneotype+Personality	Call Count, Sms_Count, Unique_location_count, InOut_ratio_Call, Time_spent_with_new_contacts, Percentage_of_time_spent_with_new_contacts, Location_Entropy, Call_Strong_ties, Sms_Strong_ties, Sms_Weak_ties, DN12am_Sms, Gyradius_Average, Returner1_vs_Explorer0, Extraversion, Agreeableness, Neuroticism, Openness	0.554***	0.00015
Phoneotype+Demographic+Personality	Call Count, Call_Response_rate, Time_spent_with_new_contacts, Percentage_of_time_spent_with_new_contacts, Inout_ratio_SMS, Location_Entropy, Call_Strong_ties, Sms_Strong_ties, Call_Weak_ties, Sms_Weak_ties, Gyradius_Average, AR_cyclicity_31, Conscientiousness, Agreeableness, Neuroticism, Openness, Gender, Family_Income	0.556***	0.00015

Adding demographic features to phoneotype features enhanced the predictive power of the model: they help increase the adjusted R-squared by a modest margin while still remaining significant. The highest adjusted R-squared score observed was 0.556 when phone-based features were combined with personality traits and demographic features. This indicates that even though demographic and personality feature alone may not be informative of one's risk propensity, when available they can be combined with phone based features for better estimates of risk propensities.

4.2 Feature Selection for Classification

A unique challenge posed in the current setting is that of the relatively large number of variables (23) given the modest sample size (N = 50). Our solution is to combine the problems of variable selection and modeling. We adopted the Recursive Feature Elimination with Linear Discriminant Analysis (RFE-LDA) similar to [30, 48] before building the predictive models based on optimal subset selection method. Starting with a set of features, RFE-LDA discards one feature that contributes the least to the performance of the LDA, and the final selected

ALGORITHM 1: Leave-one-out-nested-cross-validation on a dataset of n observations (adapted from [83])

Input: classification models, covariate data $X = \{x_i \in \mathbb{R}^m\}$, response variable $y_i \in \mathbb{R}$, $i \in \{1, 2, \cdots, n\}$.
Output: Mean accuracy of the input classification model m.
for $i = 1, 2, \cdots, n$ **do**
  Set aside data x_i as testing data;
  Do RFE-LDA on $\hat{X}^i = X \backslash \{x_i\}$, obtain dataset with selected feature $X^i_{selected}$ and x_i;
  Tune the parameter of the model on $X^i_{selected}$ using LOOCV, obtain the best model;
  Use x_i to assess the performance of the best model, obtain accuracy a_i;
  $m = \frac{1}{n} \sum_{i=1}^{n} a_i$;
end

set of features is the one that yields the best performance across different number of features. RFE allows for a heurestic based selection of a reasonable set of features when the optimal subset selection is impossible due to the number of iterations required.

4.3 Classification

There is no representative population level median data available for the survey. Hence, in order to build a classification model for risk propensities, we divided the risk scores based on median value (16) of the sample data. This resulted in two classes, which implied distinction between “high” risk taking (N=29) and “low” risk taking propensities for individuals (N=21).

Given the modest size of the dataset, and the need to maximize learning opportunities while still minimizing the odds of over-fitting, we utilized Leave-one-out-nested-cross-validation (LOONCV) for model selection and performance assessment, where for each outer iteration we used separate datasets for feature selection+parameter tuning and model assessment. Nested-cv has been shown to reduce the bias in the estimate of the testing error in prior literature [83]. A detailed description of the model selection/validation method is described below:

The described approach resulted in (n=50) different inner and outer cross-validation loops wherein different features were selected based on the abovementioned RFE-LDA approach. The most frequently selected across the 50 iterations for different models considering only the phone-based data, demography-based data and personality-based data are shown in Table 3.

Table 3. Features selected by under Nested Cross Validation across all observations

Model	Features Selected
Only Phoneotype	Call Count, SMS Count, Call response rate, Missed call percentage, SMS Strong Ties, DN12am call (selected in all 50 iterations)
Only Demographic	Gender (49 times), Race (19 times), Family Income (19 times)
Only Personality	Openness (50 times), Neuroticism (35 times)

We used the following well-established methods Logistic Regression, LDA (Linear Discriminant Analysis), GLMNET (Generalized Linear Model via lasso and elastic net regularization), Naive Bayes, Neural Network, Bayesian GLM (Generalized Linear Model), and SVM (Support Vector Machines) as the classification methods. We also report results based on a naive baseline method (Zero-R, without cross validation), which simply classifies all instances into the majority class to allow for a comparison of performance.

For the ease of presentation, we first compare the performance of the Only Phoneotype based model with Only Demographic and Only Personality models in terms of the accuracy. As can be seen in Table 4, the Only

Phoneotype approach consistently outperforms both Only Demography and Only Personality based approaches. As shown in Table 5, the phoneotype-based model yielded the best case accuracy of 78%. The only demography based model yielded a best case accuracy of 56% while the only personality had a best accuracy of 64%. Focusing on the best case results, we find that the Only Phoneotype approach yields 34% higher accuracy than the non-cross-validated Zero-R baseline and a 39% higher accuracy than the Only Demographic approach. The combination of phone-based data with personality or demographic variables did not result in any improvement in classification performance.

Table 4. Accuracy of different models at predicting Risk Propensity

Method	Only Phoneotype	Only Demographic	Only Personality
Logistic Regression	**0.78**	0.36	0.48
LDA	**0.78**	0.36	0.48
Glmnet	0.68	0.31	0.63
Naive Bayes	0.68	0.34	**0.64**
Neural Network	0.74	0.36	0.54
Bayesian GLM	0.72	0.36	0.48
SVM	0.72	**0.56**	0.54
Zero-R (baseline)	0.58	0.58	0.58

Table 5. Accuracy, Precision, Recall and F1 for different (phoneotypic) phone-based models for predicting Risk Propensity

Method	F-measure	Precision	Recall
Logistic Regression	**0.82**	**0.78**	0.86
LDA	**0.82**	**0.78**	0.86
Glmnet	0.75	0.69	0.83
Naive Bayes	0.75	0.68	0.82
Neural Network	0.78	0.77	0.79
Bayesian GLM	0.77	0.73	0.83
SVM	0.79	0.70	**0.90**
Zero-R (baseline)	0.43	0.34	0.58

To further validate and understand the classification models using phoneotype data to predict risk propensity, we expand the analysis beyond accuracy to also consider the Precision, Recall, and F-measure scores for the different algorithms. As shown in Table 5, we see a general consistency in the results across algorithms with the F-measure scores ranging between 75% to 82%. The highest Precision score obtained was 78% and the highest Recall obtained was 90%. These results suggest that phone-based features can yield models that perform quite reasonably across a number of well-established classification metrics and algorithms. Together, these results corroborate the findings from the regression analysis, that phone-based features significantly outperform demography-based features as well as personality features in prediction performance.

The limited predictive power (66% accuracy) found with personality features is somewhat surprising given past literature which suggests such an interconnection [33, 54]. We note however, that the exact interconnections between personality traits and risk propensity are far from well established. The associations reported in the literature tend to be a function of the exact personality traits inventory used, type of risk assessed, and the participant population. For instance while [54] reports a negative association between agreeableness personality trait and risky behavior, [33] reports a positive association between agreeableness and risky behavior. In fact, the

Table 6. Pearson's correlation coefficients for different phone-based features and risk-taking propensity after controlling for demographic and personality features.

Feature	**Correlation Coefficient**	**Significance Level**
Call_Count	0.008	n.s.
Sms_Count	0.145	n.s.
Unique_location_count	-0.114	n.s.
InOut_ratio_Call	-0.262	n.s.
Inout_ratio_SMS	-0.159	n.s.
Missed_call_percentage	0.240	n.s.
Call_Response_rate	0.228	n.s.
Call_Strong_Ties_Engagement_Ratio	0.132	n.s.
Sms_Strong_Ties_Engagement_Ratio	0.424	p=0.007
Call_Weak_Ties_Engagement_Ratio	-0.021	n.s.
Sms_Weak_Ties_Engagement_Ratio	-0.455	p=0.004
Call_Entropy	-0.160	n.s.
Sms_Entropy	-0.321	p=0.046
Location_Entropy	-0.035	n.s.
Time_spent_with_new_contacts	0.107	n.s.
Percentage_of_time_spent_with_new_contacts	-0.219	n.s.
Gyradius_Average	-0.330	p=0.040
Returner1_vs_Explorer0	-0.129	n.s.
DN12am_Call	0.253	n.s.
DN12am_Sms	-0.085	n.s.
Weekday_weekend_call_ratio	-0.052	n.s.
Weekday_weekend_SMS_ratio	-0.237	n.s.
Regularity_AR31	-0.084	n.s.

current work also suggests modest predictive power of personality traits on risk propensity. We do not interpret it to mean that there are no associations between personality traits and risk propensity but rather that the Big-Five personality traits may have *lesser* predictive power than phone-based features over risk propensity for college student population.

4.4 Pearson's Correlation: *Post-Hoc* Analysis

In order to investigate and further interpret the socio-mobile features considered in this work, we performed *post-hoc* Pearson's correlation tests on all the 23 features identified with respect to risk propensity score. This was undertaken using partial Pearson's correlation analysis where each correlation is controlled for the effects of demographic and personality variables. The results are shown in Table 6.

A total of four features were found to be significant. Zooming in on these features, we notice the important role played by *strong ties* in aiding one's ability to take risks. Prior literature has connected risk-taking both positively and negatively with social capital. While a negative link would suggest that individual's lacking resources (e.g., social ties and social support) would internalize the idea of taking risks to grow in their lives [49], a positive association would suggest that those with higher social capital have the social support and "cushion" to be able to take more risks in their lives [9, 12, 68].

This work suggests that the effect direction could be a function of the *type* of social capital (bridging vs. bonding) available to the individual [46, 60]. While bridging social capital (often based on weak ties) is great for access to newer information and resources, it does not provide the necessary emotional support and "cushion" in the case of a failure. Bonding social capital on the other hand builds upon the presence of strong ties and is associated strong emotional connections with a smaller set of contacts nurtured through frequent engagement [46].

The correlation analysis suggests that bonding (and not bridging) social capital may be positively associated with an individual's risk propensity. We note that the risk propensity is strongly and positively associated with the SMS strong ties engagement ratio ($r = 0.424^{**}$) and negatively associated with SMS weak ties engagement ratio ($r = -0.455^{**}$). This is further corroborated with one's risk propensity being negatively correlated with the diversity (entropy) of SMS contacts ($r = -0.311^{*}$). Lastly, the association follows a similar direction in terms of geographic interactions. Individuals with a higher risk propensity tend to focus their activities in a limited geographical radius ($r = -0.330^{**}$). Each of these suggests an association between a tendency to maintain focus and nurture a small number of ties (bonding social capital) and one's tendency to take risks.

5 DISCUSSION

5.1 Methodological Considerations

Here we discuss multiple design choices made and the corresponding caveats concerning the three types of analysis (correlation, regression, and classification) undertaken in this work.

First, we note the multiple comparisons undertaken in the correlation analysis. While such multiple comparisons are often "corrected" using Bonferroni or Bonferroni-Holm correction to maintain the confidence in the associations found, we do not do so in this work. This is because the analysis undertaken here is post-hoc and intended to help interpret the observed prediction results rather than being prescriptive in its own right. Similarly, we acknowledge the issues associated with the use of a relatively large number (23) of possibly collinear features in regression given the modest sample size (50). While this makes the interpretation of individual features difficult, the overall explanation scores of 0.499 for phoneotype (respectively 0.556 for phoneotype + demography + personality) remain interpretable. Similarly, the classification performance levels obtained (accuracy= 0.78; F-measure: 0.82) remains modest. Given the limited sample size, we focus on triangulating and identifying general trends across the three analysis methods (correlation, regression, and classification) rather than establishing hard associations between specific variables. The common question tying the three threads of analysis is: can socio-mobile signals as observed via a phone be used to infer the risk-taking propensity of an individual? The answer to this question based on the general consistency of results across the threads of analysis is positive.

5.2 Privacy of User Data and Ethical Considerations

All data used in the current study were hashed and anonymized as discussed earlier in the study design section. The exact phone numbers or the content of the calls were never available to the personnel analyzing and processing the data. The permissions required for this study (call logs) were intended to be considerably lesser than those usually required by common apps (e.g. Facebook or WhatsApp apps in Android).

We also note the moral and ethical considerations in giving a person a risk propensity based on passive data collection. While similar concerns have been raised for survey and lab-based studies with similar goals, We would suggest an explicit opt-in policy (as also adopted in this work) for any usage of this approach. While a more nuanced policy debate is required, rather than abstaining from presenting such outcomes, we choose to raise awareness about these new prospects and informing the policy debate around them.

5.3 Limitations

The current study also has a number of limitations. First, the homogeneous (Android using college students) modest sample size (50) prevents us from generalizing the results to a larger population. Hence, we would be cautious in interpreting the results until they are validated at scale with diverse populations. Next, in this study, an individual's risk propensity is based on one specific survey-based instrument. This survey instrument focuses on capturing the domain-independent, long-term risk propensity. There is literature suggesting that risk behavior is a composite of such a general propensity as well as detailed domain and situational aspects [85]. While this work only focuses on the long-term stable propensities aspect, other behavioral measurement techniques and the inclusion of domain and context may add more nuance to this understanding in future work. We also acknowledge the use of a limited set of phone based features (calls, sms, GPS) in this work. The inclusion of communication occurring via third party (e.g. WhatApp) and social media apps (e.g. Facebook) is left as part of the future work. At the same time, this work adopts the principle of "thin slices" of behavior, which suggests that even limited samples of a user's social activities may be enough to discern important characteristics about an individual [8, 81].

Lastly, all the associations reported in this work are non-causal and more detailed future experiments are needed to establish causality.

5.4 Implications

Despite these limitations, this study is the first of its kind. To our knowledge, there have been no previous studies undertaken that analyze the link between risk propensities and phone-based behavioral features. The results open the doors to a methodology that, with refinements and validation, could be used at scale. Mobile phones are now actively used by more than 3 billion users, and hence the proposed method could potentially be applied to estimate the risk propensities for billions of individuals.

In future, this work could also have multiple implications for health care professionals, health informatics professionals, mobile phone service providers, app designers, social scientists and individuals themselves. For example, an app designer could use an individual's inferred risk propensity to suggest default privacy and security related settings. Similarly, a health professional collaborating on a de-addiction app may suggest different steps to individuals based on their risk propensities. Similar settings could be suggested for mobile commerce platforms and transactions with unknown individuals in the emerging shared economy.

At the same time, the suggested methodology could help decision science researchers study risk-taking propensity at scale in the society. Besides identifying connections between spatial behavior and risk propensity, a scalable methodology to study risk could allow for asking questions regarding the spread of risk propensities in networks of billions of individuals, which are simply not possible with current survey or lab-based methods.

6 CONCLUSIONS AND FUTURE WORK

This work proposes a new methodology to automatically infer an individual's risk propensity using phone features as an alternative to classical methods like manually administered surveys. Using phone-based features allows us to build prediction models by means of machine learning classification algorithms whose performance are promising and encouraging (around 80%).

While we consider the results to be early and exploratory, they could be enhanced by future work by including a larger number of participants, more detailed phone-based features, and considering larger time durations. Furthermore, the suggested phone-based method could be expanded to study and predict other personal and societal behaviors and concepts such as gratitude, compassion, and happiness. Taken together such methods open ways to better model human beings based on ubiquitous sensing and act as a building block towards a happier, healthier society.

ACKNOWLEDGMENTS

We would like to thank Cecilia Gal, Padmapriya Subramanian, Ariana Blake, Suril Dalal, Sneha Dasari, Isha Ghosh, and Christin Jose for assisting in conducting the study and processing the data.

REFERENCES

[1] [n. d.]. Funf in a box. http://funf.org. ([n. d.]).
[2] Manuel J Sueiro Abad, Iván Sánchez-Iglesias, and Alejandra Moncayo de Tella. 2011. Evaluating risk propensity using an objective instrument. *The Spanish journal of psychology* 14, 01 (2011), 392–410.
[3] Saeed Abdullah, Mark Matthews, Elizabeth L Murnane, Geri Gay, and Tanzeem Choudhury. 2014. Towards circadian computing: early to bed and early to rise makes some of us unhealthy and sleep deprived. In *Proceedings of the 2014 ACM international joint conference on pervasive and ubiquitous computing*. ACM, 673–684.
[4] Coren L Apicella, Anna Dreber, Benjamin Campbell, Peter B Gray, Moshe Hoffman, and Anthony C Little. 2008. Testosterone and financial risk preferences. *Evolution and Human Behavior* 29, 6 (2008), 384–390.
[5] Ernest S Barratt. 1983. The biological basis of impulsiveness: the significance of timing and rhythm disorders. *Personality and individual differences* 4, 4 (1983), 387–391.
[6] Antoine Bechara. 2003. Risky Business: Emotion, Decision-Making, and Addiction. . *Journal of Gambling Studies* 19, 1 (2003), 23–51.
[7] Thilo Boeck, Jennie Fleming, and Hazel Kemshall. 2006. The context of risk decisions: does social capital make a difference?. In *Forum Qualitative Sozialforschung/Forum: Qualitative Social Research*, Vol. 7.
[8] Peter Borkenau, Nadine Mauer, Rainer Riemann, Frank M Spinath, and Alois Angleitner. 2004. Thin slices of behavior as cues of personality and intelligence. *Journal of personality and social psychology* 86, 4 (2004), 599.
[9] John Cassel. 1976. The contribution of the social environment to host resistance. *American journal of epidemiology* 104, 2 (1976), 107–123.
[10] Guillaume Chaumet, Jacques Taillard, Patricia Sagaspe, Massimo Pagani, David F Dinges, Anne Pavy-Le-Traon, Marie-Pierre Bareille, Olivier Rascol, and Pierre Philip. 2009. Confinement and sleep deprivation effects on propensity to take risks. *Aviation, space, and environmental medicine* 80, 2 (2009), 73–80.
[11] Yi-Ning Chen, Ven-Hwei Lo, Ran Wei, Xiaoge Xu, and Guoliang Zhang. 2014. A comparative study of the relationship between mobile phone use and social capital among college students in Shanghai and Taipei. *International Journal of Journalism and Mass Communication* 1 (2014), 105–113.
[12] Sidney Cobb. 1976. Social support as a moderator of life stress. *Psychosomatic medicine* 38, 5 (1976), 300–314.
[13] James S Coleman. 1988. Social capital in the creation of human capital. *American journal of sociology* 94 (1988), S95–S120.
[14] Richard A Crosby, David R Holtgrave, Ralph J DiClemente, Gina M Wingood, and Julie Ann Gayle. 2003. Social capital as a predictor of adolescents' sexual risk behavior: a state-level exploratory study. *AIDS and Behavior* 7, 3 (2003), 245–252.
[15] Per Davidsson and Benson Honig. 2003. The role of social and human capital among nascent entrepreneurs. *Journal of business venturing* 18, 3 (2003), 301–331.
[16] Yves-Alexandre de Montjoye, Jordi Quoidbach, Florent Robic, and Alex Sandy Pentland. 2013. Predicting personality using novel mobile phone-based metrics. In *International Conference on Social Computing, Behavioral-Cultural Modeling, and Prediction*. Springer, 48–55.
[17] Nathan Eagle and Alex Sandy Pentland. 2006. Reality mining: sensing complex social systems. *Personal and ubiquitous computing* 10, 4 (2006), 255–268.
[18] Tom Elfring and Willem Hulsink. 2003. Networks in entrepreneurship: The case of high-technology firms. *Small business economics* 21, 4 (2003), 409–422.
[19] Nicole B Ellison, Charles Steinfield, and Cliff Lampe. 2007. The benefits of Facebook "friends:" Social capital and college students' use of online social network sites. *Journal of Computer-Mediated Communication* 12, 4 (2007), 1143–1168.
[20] Saul Estrin, Tomasz Mickiewicz, and Ute Stephan. 2013. Entrepreneurship, social capital, and institutions: Social and commercial entrepreneurship across nations. *Entrepreneurship theory and practice* 37, 3 (2013), 479–504.
[21] Bjarke Felbo, Pål Sundsøy, Alex'Sandy' Pentland, Sune Lehmann, and Yves-Alexandre de Montjoye. 2015. Using Deep Learning to Predict Demographics from Mobile Phone Metadata. *arXiv preprint arXiv:1511.06660* (2015).
[22] Joshua Fogel and Elham Nehmad. 2009. Internet social network communities: Risk taking, trust, and privacy concerns. *Computers in human behavior* 25, 1 (2009), 153–160.
[23] Francis Fukuyama. 1995. Social capital and the global economy. *Foreign affairs* (1995), 89–103.
[24] Isha Ghosh and Vivek K Singh. 2016. Predicting Privacy Attitudes Using Phone Metadata. *In Proc. SBP-BRiMS* (2016).
[25] Eric Gilbert and Karrie Karahalios. 2009. Predicting tie strength with social media. In *Proceedings of the SIGCHI conference on human factors in computing systems*. ACM, 211–220.
[26] Jim Giles. 2012. Making the links. *Nature* 488, 7412 (2012), 448.

[27] John E Grable and So-hyun Joo. 1999. Factors related to risk tolerance: A further examination. *Consumer Interests Annual* 45 (1999), 53–58.

[28] John E Grable and So-hyun Joo. 1999. Financial help-seeking behavior: Theory and implications. *Journal of Financial Counseling and Planning* 10, 1 (1999), 14.

[29] John E Grable and So-Hyun Joo. 2000. A cross-disciplinary examination of financial risk tolerance. *Consumer Interests Annual* 46 (2000), 151–157.

[30] Pablo M Granitto, Cesare Furlanello, Franco Biasioli, and Flavia Gasperi. 2006. Recursive feature elimination with random forest for PTR-MS analysis of agroindustrial products. *Chemometrics and Intelligent Laboratory Systems* 83, 2 (2006), 83–90.

[31] Mark S Granovetter. 1973. The strength of weak ties. *American journal of sociology* (1973), 1360–1380.

[32] Mark S Granovetter. 1973. The strength of weak ties. *American journal of sociology* 78, 6 (1973), 1360–1380.

[33] Eleonora Gullone and Susan Moore. 2000. Adolescent risk-taking and the five-factor model of personality. *Journal of adolescence* 23, 4 (2000), 393–407.

[34] Christine R. Harris, Michael Jenkins, and Dale Glaser. 2006. Gender Differences in Risk Assessment: Why do Women Take Fewer Risks than Men? . *Judgment and Decision Making* 1, 1 (2006), 48.

[35] David R Holtgrave and Elke U Weber. 1993. Dimensions of risk perception for financial and health risks. *Risk analysis* 13, 5 (1993), 553–558.

[36] Hsin Hsin Chang and Su Wen Chen. 2008. The impact of online store environment cues on purchase intention: Trust and perceived risk as a mediator. *Online information review* 32, 6 (2008), 818–841.

[37] Sirkka L Jarvenpaa, Noam Tractinsky, and Lauri Saarinen. 1999. Consumer trust in an internet store: a cross-cultural validation. *Journal of Computer-Mediated Communication* 5, 2 (1999), 0–0.

[38] Oliver P John and Sanjay Srivastava. 1999. The Big Five trait taxonomy: History, measurement, and theoretical perspectives. *Handbook of personality: Theory and research* 2, 1999 (1999), 102–138.

[39] Jaemin Jung, Sylvia Chan-Olmsted, and Youngju Kim. 2013. From access to utilization: Factors affecting smartphone application use and its impacts on social and human capital acquisition in South Korea. *Journalism & Mass Communication Quarterly* 90, 4 (2013), 715–735.

[40] Daniel Kahneman and Amos Tversky. 1984. Choices, values, and frames. *American psychologist* 39, 4 (1984), 341.

[41] Vincent Kaufmann, Manfred Max Bergman, and Dominique Joye. 2004. Motility: mobility as capital. *International journal of urban and regional research* 28, 4 (2004), 745–756.

[42] William DS Killgore. 2007. Effects of sleep deprivation and morningness-eveningness traits on risk-taking. *Psychological reports* 100, 2 (2007), 613–626.

[43] Dan J Kim, Donald L Ferrin, and H Raghav Rao. 2008. A trust-based consumer decision-making model in electronic commerce: The role of trust, perceived risk, and their antecedents. *Decision support systems* 44, 2 (2008), 544–564.

[44] CW Lejuez, Will M Aklin, Heather A Jones, Jerry B Richards, David R Strong, Christopher W Kahler, and Jennifer P Read. 2003. The balloon analogue risk task (BART) differentiates smokers and nonsmokers. *Experimental and clinical psychopharmacology* 11, 1 (2003), 26.

[45] Irwin P Levin, Richard D Johnson, Patricia J Deldin, Laura M Carstens, LuAnne J Cressey, and Charles R Davis. 1986. Framing effects in decisions with completely and incompletely described alternatives. *Organizational Behavior and Human Decision Processes* 38, 1 (1986), 48–64.

[46] Nan Lin. 1999. Building a network theory of social capital. *Connections* 22, 1 (1999), 28–51.

[47] Shephard Liverant and Alvin Scodel. 1960. Internal and external control as determinants of decision making under conditions of risk. *Psychological Reports* 7, 1 (1960), 59–67.

[48] Nelmarie Louw and SJ Steel. 2006. Variable selection in kernel Fisher discriminant analysis by means of recursive feature elimination. *Computational Statistics & Data Analysis* 51, 3 (2006), 2043–2055.

[49] Natasha R Magson, Rhonda G Craven, Geoff Munns, and Alexander S Yeung. 2016. It is risky business: can social capital reduce risk-taking behaviours among disadvantaged youth? *Journal of Youth Studies* 19, 5 (2016), 569–592.

[50] Agata McCormac, Tara Zwaans, Kathryn Parsons, Dragana Calic, Marcus Butavicius, and Malcolm Pattinson. 2017. Individual differences and Information Security Awareness. *Computers in Human Behavior* 69 (2017), 151–156.

[51] Susan Moore and Eleonore Gullone. 1996. Predicting adolescent risk behavior using a personalized cost-benefit analysis. *Journal of Youth and Adolescence* 25, 3 (1996), 343–359.

[52] Reed E Nelson. 1989. The strength of strong ties: Social networks and intergroup conflict in organizations. *Academy of Management Journal* 32, 2 (1989), 377–401.

[53] Nigel Nicholson, Mark Fenton-O'Creevy, Emma Soane, and Paul Willman. 2002. Risk propensity and personality. *London Business School, Open University Business School and Said Business School Oxford, London* (2002).

[54] Nigel Nicholson, Emma Soane, Mark Fenton-O'Creevy, and Paul Willman. 2005. Personality and domain-specific risk taking. *Journal of Risk Research* 8, 2 (2005), 157–176.

[55] Nitin Nohria and Robert G Eccles. 1992. Networks and organizations: Structure, form, and action. (1992).

[56] Yue Pan and George M Zinkhan. 2006. Exploring the impact of online privacy disclosures on consumer trust. *Journal of Retailing* 82, 4 (2006), 331–338.

[57] Luca Pappalardo, Filippo Simini, Salvatore Rinzivillo, Dino Pedreschi, Fosca Giannotti, and Albert-László Barabási. 2015. Returners and explorers dichotomy in human mobility. *Nature communications* 6 (2015).

[58] Kyung-Gook Park, Sehee Han, and Lynda Lee Kaid. 2013. Does social networking service usage mediate the association between smartphone usage and social capital? *new media & society* 15, 7 (2013), 1077–1093.

[59] Robert Putnam. 2001. Social capital: Measurement and consequences. *Canadian Journal of Policy Research* 2, 1 (2001), 41–51.

[60] Robert D Putnam. 1995. Bowling alone: America's declining social capital. *Journal of democracy* 6, 1 (1995), 65–78.

[61] Owen M Quattlebaum. 1988. LOSS AVERSION: THE KEY TO DETERMINING INDIVIDUAL RISK. *Journal of Financial Planning* 1, 2 (1988).

[62] Kiran K Rachuri, Mirco Musolesi, Cecilia Mascolo, Peter J Rentfrow, Chris Longworth, and Andrius Aucinas. 2010. EmotionSense: a mobile phones based adaptive platform for experimental social psychology research. In *Proceedings of the 12th ACM international conference on Ubiquitous computing*. ACM, 281–290.

[63] Hengyi Rao, Marc Korczykowski, John Pluta, Angela Hoang, and John A Detre. 2008. Neural correlates of voluntary and involuntary risk taking in the human brain: an fMRI Study of the Balloon Analog Risk Task (BART). *Neuroimage* 42, 2 (2008), 902–910.

[64] Brian T Ratchford. 1982. Cost-benefit models for explaining consumer choice and information seeking behavior. *Management Science* 28, 2 (1982), 197–212.

[65] Elissa Redmiles, Amelia Malone, and Michelle L Mazurek. 2015. How I Learned To Be Secure: Advice Sources and Personality Factors in Cybersecurity. In *Symposium on Usable Privacy and Security*.

[66] William B Riley Jr and K Victor Chow. 1992. Asset allocation and individual risk aversion. *Financial Analysts Journal* (1992), 32–37.

[67] Rajiv N Rimal. 2001. Perceived risk and self-efficacy as motivators: Understanding individuals' long-term use of health information. *Journal of Communication* 51, 4 (2001), 633–654.

[68] María José Rodríguez-Gutiérrez, Isidoro Romero, and Zhikun Yu. 2015. The influence of social capital on risk-taking propensity. A study on Chinese immigrant entrepreneurs. *New England Journal of Entrepreneurship* 18, 2 (2015), 19–29.

[69] Michael J Roszkowski and Glenn E Snelbecker. 1990. Effects of framing on measures of risk tolerance: Financial planners are not immune. *Journal of Behavioral Economics* 19, 3 (1990), 237–246.

[70] Víctor J Rubio, JM Hernández, and J Santacreu. [n. d.]. The objective assessment of risk tendency as a personality dimension. *University Autónoma of Madrid* ([n. d.]).

[71] Sohrab Saeb, Mi Zhang, Christopher J Karr, Stephen M Schueller, Marya E Corden, Konrad P Kording, and David C Mohr. 2015. Mobile phone sensor correlates of depressive symptom severity in daily-life behavior: an exploratory study. *Journal of medical Internet research* 17, 7 (2015), e175.

[72] N Sadat Shami, Michael Muller, Aditya Pal, Mikhil Masli, and Werner Geyer. 2015. Inferring employee engagement from social media. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. ACM, 3999–4008.

[73] Claude Elwood Shannon. 2001. A mathematical theory of communication. *ACM SIGMOBILE Mobile Computing and Communications Review* 5, 1 (2001), 3–55.

[74] Vivek Singh and Isha Ghosh. 2017. Inferring Individual Social Capital Automatically via Phone Logs. *Proceedings of the ACM - Human Computer Interaction* CSCW, 2 (2017).

[75] Vivek K. Singh and Rishav R. Agarwal. 2016. Cooperative Phoneotypes: Exploring Phone-based Behavioral Markers of Cooperation. . *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. (2016), 646–657.

[76] Vivek K Singh, Laura Freeman, Bruno Lepri, and Alex Sandy Pentland. 2013. Predicting spending behavior using socio-mobile features. In *Social Computing (SocialCom), 2013 International Conference on*. IEEE, 174–179.

[77] Sim B Sitkin and Amy L Pablo. 1992. Reconceptualizing the determinants of risk behavior. *Academy of management review* 17, 1 (1992), 9–38.

[78] Sim B Sitkin and Laurie R Weingart. 1995. Determinants of risky decision-making behavior: A test of the mediating role of risk perceptions and propensity. *Academy of management Journal* 38, 6 (1995), 1573–1592.

[79] Reid L Skeel, John Neudecker, Carrie Pilarski, and Kimberley Pytlak. 2007. The utility of personality variables and behaviorally-based measures in the prediction of risk-taking behavior. *Personality and Individual Differences* 43, 1 (2007), 203–214.

[80] Charles D Spielberger. 2010. *State-Trait anxiety inventory*. Wiley Online Library.

[81] Kristin Brooke Stecher and Scott Counts. 2008. Thin Slices of Online Profile Attributes.. In *ICWSM*.

[82] Frank J Sulloway and Toni Falbo. 1997. To rebel or not to rebel? Is this the birth order question? *Psyccritiques* 42, 10 (1997), 938–939.

[83] Sudhir Varma and Richard Simon. 2006. Bias in error estimation when using cross-validation for model selection. *BMC bioinformatics* 7, 1 (2006), 91.

[84] Rui Wang, Fanglin Chen, Zhenyu Chen, Tianxing Li, Gabriella Harari, Stefanie Tignor, Xia Zhou, Dror Ben-Zeev, and Andrew T Campbell. 2014. StudentLife: assessing mental health, academic performance and behavioral trends of college students using smartphones. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, 3–14.

[85] Elke U.and Ann-Renee Blais Weber and Nancy E. Betz. 2002. A Domain-specific Risk-attitude Scale: Measuring Risk Perceptions and Risk Behaviors. *Journal of behavioral decision making* 15, 4 (2002), 263–290.
[86] Dmitri Williams. 2006. On and off the'Net: Scales for social capital in an online era. *Journal of Computer-Mediated Communication* 11, 2 (2006), 593–628.
[87] Alan Wong and Bernardo J Carducci. 1991. Sensation seeking and financial risk taking in everyday money matters. *Journal of business and psychology* 5, 4 (1991), 525–530.
[88] Ulla Y Yip. 2000. *Financial Risk Tolerance*. Ph.D. Dissertation. The University of New South Wales.
[89] Ming Yuan and Yi Lin. 2006. Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 68, 1 (2006), 49–67.
[90] Marvin Zuckerman. 1979. Sensation seeking and risk taking. In *Emotions in personality and psychopathology*. Springer, 161–197.
[91] Marvin Zuckerman. 1983. *Biological bases of sensation seeking, impulsivity, and anxiety.* Lawrence Erlbaum Assoc Inc.

Received May 2017; revised November 2017; accepted January 2018