Understanding Users' Perception Towards Automated Personality Detection with Group-specific Behavioral Data

Seoyoung Kim[1] **Arti Thakur**[2] **Juho Kim**[1]

[1]KAIST, Daejeon, Republic of Korea

[2]University of California, Davis, USA

{youthskim, juhokim}@kaist.ac.kr athakur36@gmail.com

ABSTRACT

Thanks to advanced sensing and logging technology, automatic personality assessment (APA) with users' behavioral data in the workplace is on the rise. While previous work has focused on building APA systems with high accuracy, little research has attempted to understand users' perception towards APA systems. To fill this gap, we take a mixed-methods approach: we (1) designed a survey (n=89) to understand users' social workplace behavior both online and offline and their privacy concerns; (2) built a research probe that detects personality from online and offline data streams with up to 81.3% accuracy, and deployed it for three weeks in Korea (n=32); and (3) conducted post-interviews (n=9). We identify privacy issues in sharing data and system-induced change in natural behavior as important design factors for APA systems. Our findings suggest that designers should consider the complex relationship between users' perception and system accuracy for a more user-centered APA design.

Author Keywords

User Perception; Automatic Personality Assessment (APA); Tracking; Co-located Group; Privacy; Behavior Change

CCS Concepts

•**Human-centered computing** → *Empirical studies in ubiquitous and mobile computing;*

INTRODUCTION

Personality affects one's behavior in a co-located group, where all members work in the same physical location (e.g., workplaces and university labs). Personality traits, which reflect the tendency to respond in certain ways under certain circumstances [37], significantly influence job proficiency [38], job competency [5], team formation within the group [43, 31], and social dynamics in co-located group settings [28]. Thus, it has become increasingly common to conduct personality assessment of members in co-located groups and use the results to improve group productivity. One of the biggest consumers of personality tests is organizations [10] and 88% of the Fortune 500 companies have utilized a personality test [12]. As shared by an employee of a company that asks all team members to take a personality test and shares results among the team members, knowing teammates' personalities helped resolve conflicts and enhance mutual understanding (Ally Jina Kim, personal communication, April 24, 2018).

There exist diverse methods to measure personality, each with its advantages and disadvantages. Self-assessment such as Myers-Briggs [34] and International Personality Item Pool (IPIP) [19] is widely used for high applicability, low cost of implementation, and high acceptability by users [32]. However, it requires users to spare time to take the questionnaires. Automatic personality assessment (APA) tries to address these issues by predicting the user's personality by analyzing their (1) reactivity when assigned a specific simple task or (2) everyday behavioral data. However, despite the on-going debate on existence of personality change over time, APA systems that give a specific task (e.g., giving a stimulus to track eye movements [7] or introducing oneself to capture their acoustic or visual features [6]), are one-time measurements and cannot capture personality changes over time. On the other hand, APA systems utilizing everyday behavioral data, e.g., mobile phone logs [13], social media profiles [18], or wearable device logs [35], which are collected through sensing and logging technology, have the potential to measure personality continuously without direct involvement of users. However, despite the active research on APA systems with an effort to achieve state-of-the-art performance, applying these systems in practice may face resistance from users due to the use of potentially privacy-intrusive behavioral data [39]. Without a careful understanding of users' perception towards behavioral tracking for APA, such systems would pose a threat to users, hampering them from being used in the wild.

To this end, we try to understand users' perception towards personality detecting systems with everyday behavioral data in a co-located group, which we refer to as APA hereinafter. For this, we took a mixed-methods approach: a survey with 89 full-time employees and interviews with 9 participants among 32 users who experienced our research probe, which automatically detects personality with three-week-long data collection in Korea. Our findings suggest that privacy concerns in sharing data and change in users' natural behavior induced by the system during data collection are important factors to consider while building an APA system powered by behavioral data. Lastly, we provide design implications for user-centered design of automatic personality detection

CHI'20, April 25–30, 2020, Honolulu, HI, USA

ISBN 978-1-4503-6708-0/20/04. . . $15.00

DOI: https://doi.org/10.1145/3313831.3376250

systems with behavioral data in a co-located group, emphasizing the importance of considering the complex relationship between user perception and system accuracy.

RELATED WORK

A rich body of previous work has focused on understanding users' perception such as privacy concerns, unwanted change in behavior, and trust in results towards machine learning systems utilizing their data. Many researchers have studied users' privacy concerns towards data collection, as privacy concerns affect one's mental wellbeing, productivity, and creativity [4]. For instance, users' acceptability of sharing data significantly varies between data collected from a public and private space [27]. Users may even try to avert sharing data by using backchannels with alternative instant messaging apps or social media when they had to share even personal chats from messengers or social media [2]. Further, connotations linked to data may affect willingness to share the data: users prefer sharing information with positive connotations (e.g., step counts) than negative connotations (e.g., stress levels) [40].

Another stream of work focuses on unwanted behavior change during behavioral data collection. Oftentimes, behavioral assessments are obtrusive, i.e., users become aware of the observation, which can induce reactivity, thus changing users' natural behaviors that are significantly different from their natural behaviors [25]. Behavioral data collection is not an exception; using accelerometers to measure physical activity can also cause unwanted behavior change, increasing the first few days' amount of activity [15]. Foley et al. [16] found reactivity with a pedometer as a result of providing feedback on their physical activity.

Users' trust in machine-generated decisions or information has been an active research area. Perceived accuracy in a machine learning model can be different from the real accuracy: research reveals that humans do not trust systems of which they witnessed the mistakes, despite their high accuracy, thereby causing *algorithm aversion* [14]. On the other hand, recent research suggests that users trust algorithms over humans, i.e., *algorithm appreciation*, regardless of the domain or age [30]. Yin et al. [45] found that laypeople's trust in ML models is affected by both the model's stated accuracy and its observed accuracy in practice. This highlights the importance of understanding users' perception towards ML models.

While there has been active research on automatic personality detection in recent years [7, 43], few studies have attempted to understand users' perspectives towards APA systems. Gou et al. [20] have investigated how various factors including users' own personality and perceived benefits and risks influence users' sharing preferences of derived personality traits. In addition, Warshaw et al. [42] note that users found the automatically-generated text describing their personalities creepily accurate, but would not like to share it. Likewise, previous work on understanding users of APA systems focuses on the detected personality result, rather than on how the design of APA systems can affect users' perception. Therefore, we attempt to contribute a deeper understanding of users' perception towards APA systems across various dimensions.

METHODS

To understand users' perception towards APA systems that use behavioral data in a co-located group, we took a mixed-methods approach: a survey and interviews. While our literature survey suggests that multiple factors affect people's perception (e.g., privacy concerns regarding sharing the personality result, trust in result, and system-induced change in natural behavior during data collection), a single method might not provide a comprehensive view that spans multiple factors. Through the mixed-methods approach, we attempt to combine complementary insights drawn from the different methods.

With the survey, our goal is to understand respondents' acceptability of sharing their behavioral data in diverse data streams with a specific focus on privacy. Because survey respondents often have to answer questions based on their presumption rather than actual experience, we focused on questions that could be relatively easily answered based on presumption. To further gain insights into user perception based on the actual experience of using APA systems, we built a research probe informed by the survey findings. The research probe was a custom APA system that accompanies behavioral data collection of four different data streams, which varied in the level of obtrusiveness, given user control, and technology for collection (Table 3). After 32 participants experienced the system with three weeks of data collection, we interviewed 9 participants to investigate user perception deeper.

SURVEY

We conducted an online survey (n=89) to better understand users' perception towards behavioral data sharing for personality assessment. We specifically focused on social behaviors within the context of workplace, as social behaviors commonly occur in a co-located group. We asked respondents how acceptable it is for them to share data streams within their company or organization in four aspects: (1) data collection scope across data streams (e.g., sharing online chat logs vs. offline movement logs), (2) data collection scope within a data stream (e.g., sharing online chat logs with message content vs. without message content), (3) sharing group-specific data (i.e., data that captures behaviors displayed only within a group) vs. behaviors in overall context (i.e., non-group-specific and group-specific behaviors combined), and (4) whether to have control to exclude specific data entries. Further, to understand how users' behaviors differ among the data streams, we also investigated potential differences between online and offline group behaviors, in extension to previous work [41, 26].

Differences in online and offline group social behavior patterns: We wanted to know whether there existed differences in online and offline group social behavior. We first asked how much time respondents spend on *online* and *offline* social interactions at the workplace. We chose to ask about social interaction displayed within the group as it is one of the most prevalent behaviors which can easily be found in a variety of group settings. In the survey, we explicitly gave examples of online and offline social interactions: *online* social interaction included chatting with colleagues or friends through an instant messenger and using social media, while *offline* social interaction included talking with colleagues or friends face-to-face in an informal manner, which excludes official meetings. We also asked how frequently respondents perform different social behaviors (i.e., talking a lot, starting a conversation, participating actively in a group chat, being the center of attention) *online* and *offline*.

Acceptability in sharing group-specific data with an option to exclude specific data entries: We investigated whether having an option to exclude specific data entries and sharing group-specific data (i.e., data that capture behaviors displayed only within a group) instead of sharing data in overall context can increase the acceptability in sharing data. We asked the level of acceptability in sharing certain behavioral data with their organization on a 7-point Likert scale (1-unacceptable, 7-acceptable) for each of the following three conditions of data sharing: *Data sharing condition (1)*: sharing both group-specific and non-group-specific data of a data stream, without any option to exclude data entries, *Data sharing condition (2)*: sharing group-specific data only, without any option to exclude data entries, and *Data sharing condition (3)*: sharing group-specific data only, with options to exclude specific data entries.

For the survey, we selected four online and offline data streams prevalent in modern co-located groups: online chat logs, online web or app usage logs, offline position logs, and offline movement logs.

Acceptability of sharing data across diverse data streams: We compared respondents' acceptability in sharing various data streams. We specifically focused on the data sharing level where they are asked to share group-specific data, with an option to exclude data instances (*data sharing condition (3)*). In addition to the four types of data streams (i.e. online chat logs, online web or app usage logs, offline position logs, and offline movement logs), which are easily found in modern workplaces, we also investigated their acceptability in sharing *audio* and *video* recordings, as they are richer in context yet more private. Moreover, we compared acceptability in sharing *audio* with *audio features*, such as pitch, tempo, and loudness, to reduce privacy issues within the data stream.

Acceptability of sharing data within a data stream: We wanted to understand whether sharing less information within a data stream can increase the acceptability of sharing the data stream. We asked respondents to rate the acceptability of sharing specific types of data within data streams on a 7-point Likert scale under the condition to share group-specific data, with an option to exclude unwanted data entries (*data sharing condition (3)*). Specifically, we asked in terms of three data streams as following: (1) Are people more willing to share online chat logs without message content?, (2) Are people more willing to share chat logs from public channels than private channels/DM on Slack?, (3) Are people willing to share online web/app usage data in a more abstract form (i.e., sharing URLs vs. domain information vs. categories of web/app)?, and (4) Are people more willing to share only step data than various offline movement data?

For our 50-question survey, we collected responses through various sources: an external commercial survey platform, personal contacts, and social media. As investigating differences between users with different group dynamics was beyond the scope of our research, we decided to only recruit full-time employees working in a co-located group environment to answer the survey. For quality control, we discarded respondents who spent less than 4 minutes completing our long survey (mean completion time = 22.5 minutes). From 141 initial responses, after discarding incomplete or clearly invalid answers, we ended up with a total of 89 responses (43.8% female, 33.7% aged 18∼29, 28.0% aged 30∼44, 24.7% aged 45∼60 and 13.5% aged more than 60).

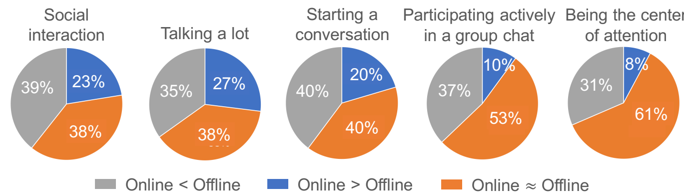

Figure 1. Online and offline social interaction patterns in the workplace: Grey represents those who show the behavior more on offline than online, blue represents those who show the behavior more on online than offline, and orange represents those who show the behavior similarly online and offline.

Result

Here we report the findings from the survey.

SR1. There exists a difference between online and offline group social behavior patterns. We categorized respondents into three groups (i.e., more online-oriented, more offline-oriented, and balanced) based on the difference between frequency or time spent for each of the social interaction *online* and *offline* as in Figure 1. Survey results show that respondents exhibit different patterns in spending time online to offline on social interaction in the workplace: 20 respondents out of 89 (22.5%) spent more time online than offline, 35 respondents (39.3%) spent more time offline than online, and the remaining 34 (38.2%) spent a similar amount of time online and offline. Moreover, the frequencies of showing each of the social behaviors in online and offline differed.

SR2. Sharing group-specific data with an option to exclude specific data entries can increase acceptability of sharing data. Respondents overall did not find it acceptable to share any of the four data streams, with average ratings between 2.15/7 and 4.4/7 for the questions asking their acceptability to share the data stream in *data sharing condition (1), (2), or (3)*. To understand how the conditions affect the acceptability of sharing data, we used Friedman's test and pairwise Wilcoxon signed-rank test with Bonferroni correction for post-hoc comparison. We observed a significant main effect of the data sharing condition on acceptability for all four data streams: online chat logs ($\chi^2(2)$=37.47), p <0.01), online web or app usage logs ($\chi^2(2)$=15.52, p <0.01), offline location logs ($\chi^2(2)$=24.17, p <0.01), and offline movement logs ($\chi^2(2)$=12.15, p <0.01). Acceptability in sharing four data streams showed a significant difference between *data sharing condition (1)* and *data sharing condition (3)* (p <0.01) as seen in Table 1. Respondents were negative about sharing non-group-specific data (*data sharing condition (1)*) (online chat logs: $M = 2.15/7$, online web/app usage logs: $M = 2.35/7$, offline location logs: $M = 2.65/7$, and offline movement logs:

	Online chat logs	Online web/app usage logs	Offline location logs	Offline movement logs
Cond. (2)−(1)	1.36**	0.67	1.00*	0.68
Cond. (3)−(2)	0.59	0.47	0.42	0.33
Cond. (3)−(1)	1.95**	1.14**	1.42**	1.01**

Table 1. Difference in acceptability between data sharing conditions for each data stream (1: unacceptable, 7: acceptable). (* p <0.05, ** p <0.01)

	Chat logs	Web/app	Location	Movement	Audio rec.	Video rec.
Web/app	*n.s.*					
Location	*n.s.*	*n.s.*				
Movement	*n.s.*	*n.s.*	*n.s.*			
Audio rec.	**	*n.s.*	**	*n.s.*		
Video rec.	**	*n.s.*	**	*n.s.*	*n.s.*	
Audio features	**	*n.s.*	*	*n.s.*	*n.s.*	*n.s.*

Table 2. Pairwise comparison of acceptability for each data stream. (* p <0.05, ** p <0.01, *n.s.*: not significant.)

$M = 2.52/7$), while they were more neutral about sharing only group-specific data with opt-out (*data sharing condition (3)*) (online chat logs: $M = 4.10/7$, online web/app usage logs: $M = 3.49/7$, offline location logs: $M = 4.07/7$, and offline movement logs: $M = 3.53/7$).

SR3. Acceptability of sharing data can differ significantly across the data streams. We found a significant main effect of the data stream on acceptability to share data (χ^2(2)=43.89, p <0.001). Unsurprisingly, respondents would likely not accept to share audio ($M = 2.82/7$) or video recordings ($M = 2.75/7$). Sharing only audio features also showed low acceptability ($M = 3.0/7$), compared to the rest of the four data streams (online chat logs: $M = 4.10/7$, online web/app usage logs: $M = 3.49/7$, offline location logs: $M = 4.07/7$ and offline movement logs: $M = 3.54/7$). While acceptability was overall low (maximum 4.1/7), our results suggest that respondents find it significantly (all p <0.05) more acceptable to share online chat logs or offline location logs compared to audio/video recording and audio features as shown in Table 2. However, for online web/app usage and offline movement, there was no significant difference in sharing compared to audio/video recording as shown in Table 2.

SR4. With a reduced scope of the data collection, acceptability of sharing data may increase acceptability in sharing. **Online chat logs.** We asked how acceptable it would be to share all chat logs *including* message content and message metadata only (e.g., timestamp, user ID, type of message (reply or not)) *excluding* message content. We did not find a significant difference between the two (p = 0.52, sharing chat log including message content: $M = 3.58/7$, sharing excluding message content: $M = 3.40/7$).

Chat logs on Slack. We asked specific questions about Slack (https://slack.com), a popular workplace online instant messenger platform. Out of 89 respondents, 22 responded they have used Slack and were qualified to answer the questions. We asked about acceptability of sharing *Direct Messages (DM)* as well as messages on *Private* and *Public* channels. Public channels differ from DM or private channels since any member in the group can access the content. From the 22 Slack users who responded, we found a significant main effect of the channel type on acceptability ((χ^2(2)=15.70, p <0.01). Post-hoc comparisons suggest that respondents are more willing to share messages from *public* channels ($M = 4.41/7$) compared to *private* channels ($M = 2.86/7$, p <0.01) and *direct messages* ($M = 2.71/7$, p <0.01).

Online web/app usage data. We asked respondents how acceptable it would be to share (1) URLs of web pages they visit or specific app activity, (2) domain information for web pages or name of the app, and (3) only categories of web page or app (e.g., social or non-social web/app). We did not find any significant differences between these levels (χ^2(2)=0.88, p = 0.65). Their acceptability in sharing online web/app usage data was all similarly low regardless of the conditions ((i): $M = 3.18/7$, (ii): $M = 3.29/7$, (ii): $M = 3.26/7$).

Offline movement data. We asked how acceptable it is to share sensor values that would indicate *movement* as well as just *steps* information. Their acceptability showed a significant difference between the two conditions (χ^2(2)=5.44, p = 0.02, any movement: $M = 3.07/7$, step: $M = 3.45/7$).

Summary of Survey Results

Our survey results suggest that analyzing both online and offline behaviors could be effective in detecting one's personality (**SR1**). Acceptability in sharing data could vary significantly depending on whether non-group-specific data is included (**SR2**), option to exclude certain data entries (**SR2**), data collection scope across the data streams (**SR3**), and even data collection scope within a data stream (**SR4**). To further understand online and offline behavior differences and privacy concerns in depth and discover additional perceptions surrounding APA, we built a research probe by applying the findings from the survey.

RESEARCH PROBE

To further understand users' perception toward APA through an actual usage experience, we built a research probe APA system. We recruited people to use the probe to understand their experience in a real context and interviewed them afterward to gain a deeper understanding of their perception. The system leverages both online and offline group-specific behaviors as people exhibit different behaviors in online to offline as found in the survey (**SR1**). It collects four different data streams (i.e., online chat logs, online web or app usage logs, offline position logs, and offline movement logs) reflecting group-specific behaviors and applying the survey results (**SR2, 3, 4**) to lower privacy concerns of participants. We also applied different levels of unobtrusiveness (e.g., giving frequent reminders throughout data collection or only in the beginning) for different data streams to better understand appropriate unobtrusiveness of an APA system. The summary of how each of the data streams is collected is shown in Table 3.

User Study with Research Probe

Our research probe involves two phases: data collection and model building. In the data collection phase, data is collected from four online and offline data streams: online messenger usage data, online web/app usage data, offline location data, and offline movement data. In the model building phase, we first extract 41 behavior features shown in Table 4. Then, the features are fed to a machine learning model to classify a user into one of the three classes (i.e. low, medium, and high) for each of the Big Five Personality traits [19].

We recruited four different research groups in the college of engineering at a large technical university in Korea to use our research probe. Some members of these groups chose not to participate in the study due to their own reasons. Excluding them, the four groups consisted of five, seven, nine, and eleven participants respectively, for a total of 32 participants (19% female, mean age = 26.7, S.D. = 3.7, 87.5% Korean). Each group used a single shared space, without individual offices. The four groups varied in their culture, social dynamics, and space utilization: (1) while two groups use Korean's honorific language to communicate with each other, other groups would use it to only those who are older, (2) in one group people more closely work with each other within internal teams, while in the other groups people rather work independently on their own projects, and (3) two groups have a common area for informal social interaction while the other groups do not. Each participant received $30 for their participation in a three-week long data collection for the research probe. Institutional Review Board (IRB) approval was obtained from the university prior to the study. Participants were also asked to read and sign the terms of use, which contained information about the purpose of the study and the scope of the data collection along with study guidelines.

Phase 1: Collecting Behavioral Data

For data collection, participants were provided with a smartwatch (ASUS ZenWatch 3) and instructed to charge it whenever required and wear it. For participants who were not actively logging for several days, researchers reminded them. During three weeks in May 2018, we collected an average of 47.8 hours of offline location and movement data per person, total 2,690 online messenger activity logs (e.g., chats, reactions, participants leaving or joining a channel), and an average of 27.0 hours of online web/app usage data per person.

For the sake of transparency, after the data collection, we asked each participant to retrieve and review their online data before sharing it with us. We then provided each individual with a summary of their data as in Figure 2. Before analyzing the data, we gave them an option to exclude any data they want. None excluded any data instance.

To collect ground truth personality data, we asked each participant to take the International Personality Item Pool (IPIP) [19] with 100 short questions to be answered in the context of their lab to measure five dimensions of personality traits (i.e. openness, conscientiousness, extraversion, agreeableness, neuroticism) [10]. With the questionnaire result, we classified participants into three levels of each personality trait by defining the middle class as those defined with scores within one SD from the mean. This is done as the small score difference within the same class could be due to report bias. Although ground truth personality was measured six months after the data collection, it was reliable, as it showed 84.3% concurrence (27 out of 32 participants, with no participant having dramatic change of personality class—introvert changed to extrovert or vice versa) with the extraversion result which was measured right after the data collection. This is consistent

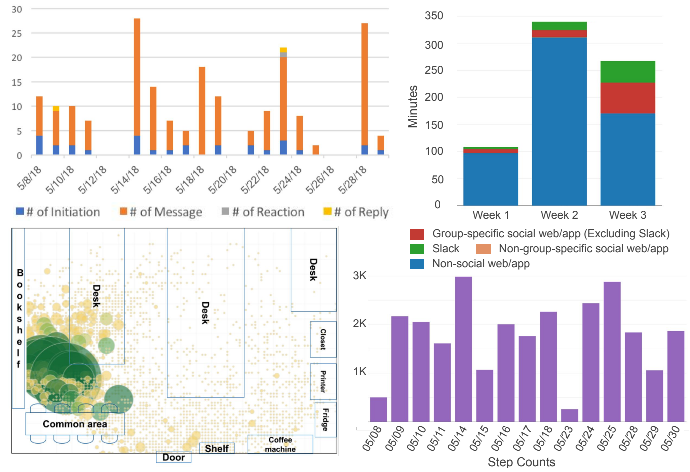

Figure 2. Summary diagrams provided to the participants after the data collection. From the top left, each summary diagram represents summary of collected messenger usage data, web/app usage data, location data, and movement data.

with previous work that personality is relatively stable over time [11].

Online Messenger Usage Data Messenger logs contain various clues to infer one's personality [29, 23, 44], making it an appropriate source of data stream for an APA system. At the same time, it could be considered as an intrusive data stream with personal chat history. Therefore, to be less invasive, we analyze only the timestamps of messages and logs of messenger activities (e.g., message, reply, and reaction) without text content (**SR4**). We collected group messenger logs (**SR2**) of Slack, which was the communication app used in all four co-located groups that we collected data from. From the logs, we only collected public channel logs as private channels and direct messages tend to be more personal as shown in the survey (**SR2**). We also gave participants the control to discard some logs before they share the data (**SR2**). We collected messenger logs with the lowest intended level of obtrusiveness, by informing what kind of data is going to be collected only at the beginning of the study.

Online Web/app Usage Data Online web/app usage data represents digital traces of a person's online behaviors. However, collecting all raw traces could lead to privacy issues. Therefore, we confined the collected data to when the person was physically in the co-located group space (**SR2**) on the weekdays. Furthermore, we provided additional control options to participants to stop logging for a certain amount of time and to exclude some of the data instances (**SR2**). We collected the web/app usage data using RescueTime (`https://www.rescuetime.com`). Even though the survey result shows no significant difference in sharing different levels of web/app usage data, we collected domain information only for Slack. For other web or apps, we only categorized them as group-specific-social, non-group-specific-social, or non-social web/app usages (**SR4**). To discern social versus non-social web/app, we followed the classification provided by RescueTime. To discern group-specific versus non-group-specific web/app, we asked participants to fill a survey to identify whether a certain social web/app is used to interact within the group. We collected web/app usage data with a relatively

Type of data	Collected content	Collected time	Data source	Intended level of obtrusiveness	Given user control	Tools used for collection
Online Messenger Usage Data	Public channel logs excluding text content	At all times	Group messenger	Low (*Only informed before the data collection*)	Can exclude data after the data collection	Slack
Online Web/App Usage Data	Timestamp of access, time spent, category of web/app	Weekdays	Inside the co-located space only	High (*Informed before the data collection, real-time/weekly report on collected data*)	Can turn off RescueTime, can exclude data after the data collection	RescueTime
Offline Location Data	Timestamp, location inside the lab in (x, y)	Weekdays	Inside the co-located space only	Medium (*Informed before the data collection, constantly giving task to make aware of data collection*)	Can turn off the watch, can exclude data after the data collection	BLE beacons, smartwatch
Offline Movement Data	Step counts, timestamp of detected step	Weekdays	Inside the co-located space only	Medium (*Informed before the data collection, constantly giving task to make aware of data collection*)	Can turn off the watch, can exclude data after the data collection	Smartwatch

Table 3. Summary of collected data.

high intended level of obtrusiveness: participants were continuously made aware of their online web/app usage tracking through a real-time dashboard and weekly reports provided by RescueTime.

Offline Location Data Location traces of a person can be used to infer one's personality. For example, it is shown that one's GPS logs of everyday life correlate with personality [13]. We collected participants' location information inside a physical co-located group space during weekdays (**SR2**). To collect offline location data, we developed an Android app for wearable devices that receives signals from BLE beacons (Estimote Location Beacons (`https://estimote.com/products`)) installed around the walls of each group's space and calculates the user's indoor position inside the space. We deployed the app in an off-the-shelf wearable smartwatch. Participants could pause data collection by turning off our data collection app or turning off the smartwatch and could exclude data instances at the end of data collection (**SR2**). We collected offline location data with a mid-level of unobtrusiveness: participants were constantly aware of the location data collection as they had to wear and charge the watch and were reminded to turn it on.

Offline Movement Data Movement inside a co-located group space can be indicative of one's personality trait [1, 21]. We collected movement information only within the co-located group space and excluded data over weekends to lessen the privacy concerns (**SR2**). Moreover, instead of collecting data regarding various activity information (e.g., walking, running, sitting) which can be considered intrusive by users, we only collected step count information as it is elementary information required to detect agility (**SR4**). We developed the app installed in the smartwatch to collect step counts and timestamps for each step. Participants could stop logging their movement and exclude data instances at the end of the study (**SR2**). Data was collected with a mid-intended-level of obtrusiveness with the same measure as offline location data collection: they had to wear and charge the watch and were reminded to turn it on.

Phase 2: Building an APA Model

From each of the collected data streams, we extracted 41 behavior features (19 online features and 22 offline features) as shown in Table 4. The full list of extracted features and how they are extracted can be found in the supplementary material. We then post-processed all behavior features to minimize the effect of differences in the group culture. For example, one group had lots of reactions added to others' messages in online messenger logs, while another group barely had any reactions. Individuals might adhere to their group culture irrespective of their own personality traits. To prevent each group's custom from influencing users' detected personality, we standardized each user behavior relative to one's own group so that every behavior feature in each group has a mean of 0 and a standard deviation of 1.

With the processed behavior features, we built an APA model to determine each participant's level of each of the Big Five personality dimensions. We ran Leave-One-Out Cross-Validation to prevent overfitting and oversampled small-numbered-classes to balance out the classes using a variant of SMOTE algorithm. Note that we performed oversampling using only the training dataset for every 10-fold cross-validation. We selected features to prevent overfitting before oversampling the training set. Then we selected the best model for each personality among a range of classification algorithms (Linear SVC, Gaussian Process Classifier, Decision Tree Classifier, Random Forest Classifier, and Gaussian NB) based on not only the high accuracy but also on F_1 macro score.

Result of the APA Model

The best model prediction accuracy for each of the Big Five Personality traits is as following: 81.3% for openness (F_1 macro score: 71.4%), 75.0% for conscientiousness (F_1 macro score: 65.8%), 81.3% for extraversion (F_1 macro score: 46.4%), 81.3% for agreeableness (F_1 macro score: 60.5%), and 71.9% for neuroticism (F_1 macro score: 58.1%). Even though the performance of our APA model suggests one possible design of automatic personality assessment with behavioral

Online	Offline
	• Staying / not staying at one's seat (**C**, **E**, **N**)
• Using (1) social-related web/app, (2) group-specific-social web/app, and (3) Slack (**E**, **N**)	• Staying in a common area (**O**)
• Accessing (1) social-related web/app and (2) group-specific-social web/app (**O**, **A**)	• Going to common area (**C**)
• Initiating a conversation on (1) any Slack channel and (2) only Slack channels including everyone	• Staying together with other group members at other than one's own seat (**C**)
• Sending / not sending a text message on (1) any Slack channel, (2) only Slack channels including everyone (**A**, **N**)	• Staying together with other group members in a common area
• Replying to others on (1) any Slack channel, (2) only Slack channels including everyone	• Arriving at the lab
• Reacting to others on (1) any Slack channels, (2) only Slack channels including everyone (**N**)	• Walking / not walking (**E**)

Table 4. List of group-specific behaviors that were analyzed for predicting personality. The bold alphabets represent the personality traits (O: Openness, C: Conscientiousness, E: Extraversion, A: Agreeableness, and N: Neuroticism) that selected the corresponding behavior as one of the top three frequently selected features throughout the cross-validation folds.

data in a co-located group, we note that we do not state the APA model itself as a research contribution in this paper, as the number of participants that we recruited is not big enough to verify its performance; we only use the result of the probe to create a realistic usage experience of an APA system to collect richer insights through the interview.

INTERVIEW

To understand users' perspective towards our research probe, we recruited 9 out of 32 participants (at least one participant from each of the four groups) who participated in the study for a post-interview. We conducted the interview in a semi-structured format for around 30 minutes and participants received $10 for compensation. During the interview, we showed them their collected data and asked their opinions regarding sharing the data for automatic personality assessment in the group. Moreover, we asked their opinion on sharing their personality result driven from their behavioral data while showing them the system prediction and their self-assessed result. We specifically focused on extraversion for this, as extraversion was a relatively widely known personality trait even among laypeople. Moreover, our focus was on sharing the detected personality from their behavioral data, not on investigating the differences in willingness to share different personality traits which were already investigated by Gou et al. [20]. Because the prediction accuracy could affect participants' perception of and trust in the system [45], we intentionally recruited both participants whose personality was predicted correctly by the system (n=5) and those with incorrect prediction (n=4).

To analyze the interview data, we transcribed the interview recordings and conducted a thematic analysis. Thematic analysis is a widely used qualitative research method to identify salient patterns or themes in the data [9]. We conducted a thematic analysis in the following five phases: (1) read transcripts while making notes, (2) go over the notes and categorize the notes, (3) tag and label themes, (4) revise tags and themes twice, and (5) re-examine tags. The first three phases were conducted by three researchers; (4) were done by two researchers; (5) was done by one researcher.

Result

The set of themes and codes resulting from our thematic analysis are presented in Table 5. In this section, we introduce each code in detail. In the rest of the paper, we use the shorthand codes presented in Table 5 to reference the codes (e.g., **BD** indicates potential benefits of using data).

Theme 1. Were participants concerned about privacy? Overall, participants did not express much privacy concerns in the first place, even though some would state some concerns as we asked specifically. Participants' perception on privacy varied due to various factors: scope and nature of the data, benefits provided, transparency, social desirability, and control over data collection. Some participants expressed concerns about sharing their data or personality prediction result due to: (1) potential misuse of the data other than the original purpose of data collection or intuitive concerns without a clear reason (**PY1**), and (2) clear privacy concerns towards the current scope/purpose of data collection (**PY2**). Examples of potential misuse (**PY1**) that participants mentioned were surveillance, regulating work styles, and assigning high-level meanings to data (e.g., assuming that a high step count means you worked hard). Even though they were notified of the purpose of collecting data, they would still be concerned due to possible misuse cases they could imagine on. Moreover, some expressed concerns but could not address a clear reason (**PY1**). On the other hand, participants who were concerned with a clear reason (**PY2**) pointed out different acceptability towards sharing different data streams (**SR3**): ("*(web and app were more intrusive than other data streams) because it's more personal rather than public*" (P25). Factors that affected the level of privacy concern towards a data stream were whether there exists social desirability in the data stream and whether the data stream is capturing personal behaviors rather than group-specific behaviors (**SR2**). Moreover, some expressed different levels of concerns even within the same data stream: "*not collecting the timestamp of web/app usage would alleviate the privacy issues a lot, rather collecting duration (of each web/app usage instance) would be better*" (P32). This implies that users' privacy concerns may be relieved with additional filtering within the data stream (**SR4**).

On the other hand, participants who reported no privacy concern gave the following reasons: (1) the characteristic of data or personality is limited (e.g., specific information within the data stream not being collected or being abstract (**SR4**), group-specific characteristic (**SR2**)), (2) a direct interpretation of data or personality prediction results is difficult due to the partial

Theme		Code	Example
Were participants concerned about **privacy**?	Yes	Due to **potential/imaginary** misuse (**PY1**)	"*If it is used for surveillance purpose, it will definitely be uncomfortable regardless of what. Even with the meaningless data.*" (P27)
		Due to **clear** reason (**PY2**)	"*It will feel a bit awkward to show both the part of me that I want to show and I don't want to show (if I share my personality with others.*" (P20)
	No	Due to the **characteristic** of data/personality trait (**PN1**)	"*I feel sharing step data and messenger data are not intrusive) because . . . if I'm walking then everybody in the lab is seeing that I'm walking. And if I'm chatting in a public channel, everybody can see that I'm chatting in a public channel, everybody can see that I'm doing that*" (P25)
		Due to the **representation** of data/personality result (**PN2**)	"*(I didn't want to erase any web/app usage data) because it doesn't really show you much. It's too abstract. You cannot know whether I talked to someone through Facebook messenger or whether I looked at certain page. . .* " (P27)
		Due to other reasons (**PN3**)	"*But this kind of thing where I can collect data myself and then I can see before I share it. . . it is really important. And you can have much more trust.*" (P22)
What affects participants' **behaviors change?**		Observer effect (**CO**)	"*(My web/app usage behaviors) could have been a bit different from when I was logging to when I was not.*" (P11)
		Self recognition (**CR**)	"*I didn't change my behavior on Slack because I couldn't think I was being tracked.*" (P22)
What affects participants' **trust** in personality result?		Data-driven aspect of the result (**TD**)	" *I was worried that I got introvert for not wearing watch*" (P28)
		Self-perception of their own personality (**TP**)	"*I thought I was mostly quiet in the lab. It was surprising that the system predicted that I was an extrovert. Basically it didn't make sense to me*" (P30)
		Ambiguity around system (**TS**)	"*I was a bit curious whether the system predicted introverts based on online and offline data.*" (P30))
What are the potential **benefits**?		Using data (**BD**)	"*I know that that (movement) data is potentially useful for me because it seems that I'm not moving around, I should move more.*" (P22)
		Knowing personality (**BP**)	"*Data could be very useful if I want to contact somebody. So if there's like 5 React experts in the lab, and 4 of them are introverts and 1 of them is extrovert, then I'd be more likely to ask the extrovert first than the introvert,. . .* " (P30)
		Regarding system (**BS**)	"*If I do data collection in a long term, it is more objective (than traditional personality test). . . Also I think it is more convenient as I don't have to think (to answer to traditional personality test).*" (P29)

Table 5. Results of our thematic analysis show four emerging themes: (1) privacy, (2) behavior change, (3) trust in result, and (4) benefits.

information that is logged instead of in a video format which shows what you did directly and the format of saved data (e.g., locations are saved in coordinates instead of a dot on a floor map), and (3) other reasons (e.g., transparent data transferring process (i.e. participants retrieving their own data to us after reviewing/deleting some unwanted instances to share for online data), trust in who they are sharing with, imperfection in data, agreement about data collection made beforehand).

Theme 2. Did the system induce participants' behavior change during data collection? Participants reported mixed responses when asked whether their behavior changed due to the system during the data collection phase. Participants who reported system-induced change of behavior said they changed their web/app usage but not others, which were deployed with high intended level of obtrusiveness as in Table 3. They pointed out privacy concerns arising from the observer effect (**CO**) and self-monitoring (**CR**). The observer effect refers to the unwanted change in behavior of the subject under observation due to the awareness of being observed [22]. For instance, P33 said, *"There was a feeling that someone was watching me and my behavior seems to change because of that.*" (**CR**). P29 said, *"(As I'm using web/app tracker,) I could track my web/app usage, so I wasted time (on my computer) less than usual."* (**CR**). However, many participants with behavior change from self-monitoring also reported that their behavior returned gradually. This aligns with P4's statement, who used to react to the daily summary provided as a long-time user of RescueTime, but after using it for a while, he does not anymore. Interestingly, several participants specifically pointed out that they did not change messenger usage behavior, which was collected with low obtrusiveness in contrast with web/app usage behavior. From this, we could know obtrusiveness of data collection can induce unwanted change in users' natural behaviors.

Theme 3. What affects participants' trust in personality result? Participants reported that their trust in results was affected by the following factors: (1) data-driven aspect of the system (**TD**), (2) self-perception of their own personality (**TP**), and (3) ambiguity around how the system works (**TS**). The data-driven aspect of the system had a mixed effect on participants' trust in the personality prediction result. For instance, P25 said, *"(The result is based on) three weeks (of data). It's longer time... and it's a data-driven approach. So, I think your system is quite accurate."* In fact, for some participants, even though they were provided with predicted personality which was different from their self-assessed personality, they showed trust towards the system-driven personality. On the other hand, some participants displayed a sign of disbelief due to the limited data collection scope when presented with their result: *"I used (Facebook) through my phone and then I didn't opt to track my phone so... (I'm not sure whether enough of my behaviors are captured by the system)."* (P22). This indicates that utilization of data affects participants' acceptability towards the results. Participants' self-perceived personality also affected trust in predictions. Even though P20 and P30 both had prediction result different from their self-assessment, P20 stated, *"(The predicted) Self-assessment is as expected"*. On the contrary, P30 questioned the result and said: *"I thought I was mostly quiet in the lab. It was surprising that the system predicted that I was an extrovert. Basically, it didn't make sense to me"*. Regardless of the actual accuracy, congruence between the prediction with their self-perceived personality would affect their trust in results. In addition, some participants blamed lack of transparency on how the system works to predict personality as reasons for their distrust as in Table 5 (**TS**). Thus, providing the reason behind the prediction could lower their distrust.

Theme 4. What are the potential benefits? Several participants mentioned the possible benefits they could receive from (1) utilizing data itself for other purposes (e.g., reflecting/recalling on oneself's productivity, interacting with others by sharing the data, space planning based on location data) (**BD**), (2) knowing their own/others' personality (e.g., being more confident about oneself, good for new-comers to know other members, asking questions to extroverts easily) (**BP**), and (3) using APA system utilizing behavioral data over other personality measurements (**BS**). Participants reported they prefer the APA system over traditional self-assessment (**BS**) due to the convenience of easily knowing personality and 'objectiveness' in the result: *"When assessing myself, today I might feel cheerful that I might answer that I'm more sociable, but tomorrow I might be depressed. So I don't really trust it because it can result differently every day. So if data is collected for a long time and analyzed, I think that personality is more reliable. And it was more convenient that I didn't have to think a lot."* (P20). This highlights the benefit of assessing personality with users' natural behavioral data.

DISCUSSION

In this section, we discuss design implications for user-centered design of an APA system: accuracy, privacy concerns, and system-induced change in users' natural behavior. Then we discuss the limitations of our study.

Implications for User-centered APA Design

Considering users' privacy concerns. In order for APA to be used in the wild, users' perception of privacy should be carefully considered. Sources of users' privacy concerns could be classified into two: (1) potential misuse or intuitive discomfort without a clear reason (**PY1**), and (2) rational thinking around current scope/purpose of data collection (**PY2**). We address possible ways to alleviate each concern. From the interview, we found that the first type of concern (**PY1**) can be alleviated by showing users the raw data to relieve users' anxiety. After showing the collected data, P29 said, "*Actually, there are much fewer privacy issues than I originally thought, as the content of the messages is all erased...*" This aligns with previous work on the privacy paradox [36]. Although we showed detailed terms of use—including what data is specifically collected and how—and a high-level individual summary of the collected data, they still gave users the room for imagination on what is collected. This raises *intuitive concerns* [36], although their *considered concerns* could be smaller with rational thinking on the actual scope. Thus, to minimize *intuitive concern* [36] (**PY1**), in addition to providing a high-level summary of data collection scope, providing raw data with an appropriate explanation of its use could alleviate concerns. In addition, trust in the person they are sharing the data/personality result with (**PN3**) also plays a pivotal role in users' privacy perception. P27 said, "*I don't like to share location data (with the professor) if they care whether I move around a lot.*" If users do not have enough trust in the person they are sharing with and think that they will use the data otherwise, their concerns on misuse would persist.

Another type of privacy concern arising from users' rational thinking around current scope/purpose of data collection (**PY2**), could be eased by taking preventive measures while designing an APA system. First, the characteristics of the data streams to be analyzed should be considered. Data streams that capture behaviors with clear social desirability or personal behaviors should be refrained from selection. Second, the scope of data collection even within the same data stream should be taken into account. For instance, our survey and interview results collecting only group-specific behavioral data (**SR2**) or reducing the scope of data collection within the stream to remove intrusive elements (**SR4**) could lower users' privacy concerns. Moreover, collecting the data in abstract form would be better so that raw data limits direct interpretation about the person. For example, many participants stated that their concern levels differ within the web/app data usage by the inclusion of the exact domain addresses they visited. Lastly, the measures taken to collect data also greatly influence users' privacy concerns. Giving users an option to review and exclude collected data or control to pause data collection can relieve their concern of constantly being tracked. Furthermore, the choice of technology used for the data collection could affect user's acceptability towards sharing data. For example, several participants reported that inaccuracy in offline position data alleviated their privacy concerns due to the uncertainty that is present for others to interpret one's exact position from the collected data. Participants reported that using smartwatches to share offline position information, which can be freely turned off, and beacons, which has approximately a $1m$ of error, helped reduce their perceived privacy concerns.

Moreover, users' privacy concerns regarding sharing personality results should be considered, where the factors influencing their level of privacy concerns are similar to as when they are sharing their behavioral data. Social desirability towards the same personality could be different among users: some users may believe that being more of an extrovert is less desired in workplace settings and others may think the opposite, as we found in the interview. Therefore, it is important to take into account the group culture and their interpretation of the personality traits in their unique settings.

Considering system-induced change in users' natural behavior. Unwanted change in users' natural behavior induced by the system can affect the accuracy of the system. Moreover, as users are aware of their own behavior change, their trust towards the system result can be also influenced (**TD**). System-induced change in users' behaviors is inevitable even without direct elicitation from users (e.g., giving users a task such as short presentation [6]). However, unobtrusive measures, i.e., completely not informing users, to eliminate reactivity in natural behavior can be unethical due to privacy intrusion. According to the interview, users' reported reactivity during the study was due to (1) privacy concerns arising from the observer effect (**CO**) and (2) self-monitoring caused by the high level of obtrusiveness (**CR**). It is hard to recover users' natural behaviors if the reason behind their behavior change is privacy-related; none of the participants whose behavior changed due to the observer effect reported any sign of recovering their natural behaviors. On the other hand, if the reason is due to self-monitoring caused by the obtrusiveness of the system, users' behavior is likely to return to their natural state after a while as they get accustomed to it, which aligns with findings on reactivity [24]. Therefore, for long-term deployments of an APA system, we emphasize the importance of considering users' privacy concerns to minimize unwanted behavior change. If the data stream to be collected is privacy-wise intrusive, even keeping a high level of obtrusiveness (e.g., providing raw data or a summary of collected logs periodically) to lower privacy concerns is suggested.

Data streams to collect for better personality detection. For better personality prediction, it is important to analyze various data streams. This is because as behaviors are responses to trait-relevant situational cues [33, 8], behavior expressions may vary across various data streams due to different degrees of situational impact in these data streams. As our survey (**SR1**) suggests, one way to capture diverse behaviors is to analyze both online and offline behaviors. This can be also seen from the results of our research probe deployment: behaviors that were related to the top three selected features for the best models differed between traits as in Table 4. While the model for conscientiousness mostly used offline behavior features, agreeableness used mostly online behavior features, and for openness, extraversion, and neuroticism, both online and offline features were used. Hence, *analyzing both online and offline behavioral data can result in better personality detection*.

Complex relationship between accuracy and users' perception. Although it may require compromise in accuracy, it is not suggested to exhaustively collect user's data just focusing on the system accuracy. The data streams we considered in this work do not cover all trait-relevant user behaviors. Previous research suggests that minute behaviors such as voice tone [17], hand movement, and posture [3] can be useful in predicting personality. While these minute behavior could have been captured using audio/video recordings, designers of future APA systems should be careful in including these data streams, as our survey results indicated respondents' reluctance to share audio/video data (**SR3**). Respondents even reported low acceptability for sharing only certain audio features such as pitch, tempo, and loudness. Moreover, it is important to note that users' trust in results is not solely determined by the accuracy of the prediction model but also through the data-driven aspect of the system (**TD**), self-perception on their own personality (**TP**), and transparency of the prediction mechanism (**TS**). If the user changes behavior due to privacy concerns aroused from excessive data collection, their perceived accuracy could degrade. Therefore, it is rather desirable to consider the gain in system accuracy relative to the cost in users' perception for the system to be actually used in the wild. Although consideration of user perception may cause degradation in system accuracy, it should not be ignored for the successful deployment of the system.

Limitations

There are several limitations of our study. First, the deployment of our research probe was done in four academic research groups in Korea. As mentioned in Section 5.1, even though the four groups had different their culture, social dynamics, and space utilization, participants' perception might only represent users in academic environments or the Asian culture. Second, we interviewed participants by showing the predicted extraversion trained with the self-assessment result that was measured right after the data collection. Although it is different from the result shown in Section 5.1.3, it has 84.3% concurrence with no dramatic class differences. Moreover, our purpose was to understand the users' acceptability of sharing the predicted result and trust towards it for both correct and incorrect personality results as every system may have failure cases. Lastly, we used a mix of recruiting methods for the survey, which could lead to generalizability issue.

CONCLUSION

We investigated users' perception towards automatic personality assessment (APA) through a mixed-methods approach: a survey and interviews with participants after experiencing our research probe. We present design implications that highlight the importance of considering users' privacy concerns and system-induced change in natural behavior for designing APA systems using behavioral data in the wild. We believe that our work opens doors for more user-centered APA design to be used in the wild.

ACKNOWLEDGEMENTS

This work was supported by Institute of Information & Communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (No.2016-0-00564, Development of Intelligent Interaction Technology Based on Context Awareness and Human Intention Understanding and No.2017-0-00537, Development of Autonomous IoT Collaboration Framework for Space Intelligence).

REFERENCES

[1] Mark S Allen, Emma E Walter, and Máirtín S McDermott. 2017. Personality and sedentary behavior: A systematic review and meta-analysis. *Health Psychology* 36, 3 (2017), 255.

[2] Lyndsey L Bakewell, Konstantina Vasileiou, Kiel S Long, Mark Atkinson, Helen Rice, Manuela Barreto, Julie Barnett, Michael Wilson, Shaun Lawson, and John Vines. 2018. Everything we do, everything we press: Data-driven remote performance management in a mobile workplace. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, 371.

[3] Gene Ball and Jack Breese. 2000. Relating personality and behavior: posture and gestures. In *Affective interactions*. Springer, 196–203.

[4] Kirstie Ball. 2010. Workplace surveillance: An overview. *Labor History* 51, 1 (2010), 87–106.

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

[6] Ligia Maria Batrinca, Nadia Mana, Bruno Lepri, Fabio Pianesi, and Nicu Sebe. 2011. Please, tell me about yourself: automatic personality assessment using short self-presentations. In *Proceedings of the 13th international conference on multimodal interfaces*. ACM, 255–262.

[7] Shlomo Berkovsky, Ronnie Taib, Irena Koprinska, Eileen Wang, Yucheng Zeng, Jingjie Li, and Sabina Kleitman. 2019. Detecting Personality Traits Using Eye-Tracking Data. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, 221.

[8] Tim Blumer and Nicola Döring. 2012. Are we the same online? The expression of the five factor personality traits on the computer and the Internet. *Cyberpsychology: Journal of Psychosocial Research on Cyberspace* 6, 3 (2012).

[9] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative research in psychology* 3, 2 (2006), 77–101.

[10] Neil Christiansen and Robert Tett. 2013. *Handbook of personality at work*. Routledge.

[11] Philip J Corr and Gerald Matthews. 2009. *The Cambridge handbook of personality psychology*. Cambridge University Press Cambridge, UK:.

[12] CPP. 2018. CPP — The MyersBriggs®Company. (2018). `https://www.cpp.com/` Last Accessed: 2018-08-09.

[13] 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.

[14] Berkeley J Dietvorst, Joseph P Simmons, and Cade Massey. 2015. Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General* 144, 1 (2015), 114.

[15] Alain Dössegger, Nicole Ruch, Gerda Jimmy, Charlotte Braun-Fahrländer, Urs Mäder, Johanna Hänggi, Heidi Hofmann, Jardena J Puder, Susi Kriemler, and Bettina Bringolf-Isler. 2014. Reactivity to accelerometer measurement of children and adolescents. *Medicine and science in sports and exercise* 46, 6 (2014), 1140.

[16] John T Foley, Michael W Beets, and Bradley J Cardinal. 2011. Monitoring children's physical activity with pedometers: Reactivity revisited. *Journal of Exercise Science & Fitness* 9, 2 (2011), 82–86.

[17] Howard S Friedman, M Robin DiMatteo, and Angelo Taranta. 1980. A study of the relationship between individual differences in nonverbal expressiveness and factors of personality and social interaction. *Journal of Research in Personality* 14, 3 (1980), 351–364.

[18] Jennifer Golbeck, Cristina Robles, and Karen Turner. 2011. Predicting personality with social media. In *CHI'11 extended abstracts on human factors in computing systems*. ACM, 253–262.

[19] Lewis R Goldberg. 1992. The development of markers for the Big-Five factor structure. *Psychological assessment* 4, 1 (1992), 26.

[20] Liang Gou, Michelle X Zhou, and Huahai Yang. 2014. KnowMe and ShareMe: understanding automatically discovered personality traits from social media and user sharing preferences. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 955–964.

[21] Heather Hausenblas and Ryan E Rhodes. 2016. ExErcisE Psychology. (2016).

[22] Paul P Heppner, Bruce E Wampold, and Dennis M Kivlighan. 2007. Research design in counseling: Research, statistics, & program evaluation. *Cengage Learning* (2007).

[23] Everard Ho and Vichita Vathanophas. 2003. Relating personality traits and prior knowledge to focus group process and outcome: an exploratory research. *PACIS 2003 Proceedings* (2003), 67.

[24] Alan E Kazdin. 1974. Reactive self-monitoring: the effects of response desirability, goal setting, and feedback. *Journal of consulting and clinical psychology* 42, 5 (1974), 704.

[25] Alan E Kazdin. 1979. Unobtrusive measures in behavioral assessment. *Journal of Applied Behavior Analysis* 12, 4 (1979), 713–724.

[26] Seoyoung Kim, Jiyoun Ha, and Juho Kim. 2018. Detecting Personality Unobtrusively from Users' Online and Offline Workplace Behaviors. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, LBW515.

[27] Sandjar Kozubaev, Fernando Rochaix, Carl DiSalvo, and Christopher A Le Dantec. 2019. Spaces and Traces: Implications of Smart Technology in Public Housing. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, 439.

[28] Kibeom Lee, Michael C Ashton, and Kang-Hyun Shin. 2005. Personality correlates of workplace anti-social behavior. *Applied Psychology* 54, 1 (2005), 81–98.

[29] Weijian Li, Yuxiao Chen, Tianran Hu, and Jiebo Luo. 2018. Mining the Relationship between Emoji Usage Patterns and Personality. *arXiv preprint arXiv:1804.05143* (2018).

[30] Jennifer M Logg, Julia A Minson, and Don A Moore. 2019. Algorithm appreciation: People prefer algorithmic to human judgment. *Organizational Behavior and Human Decision Processes* 151 (2019), 90–103.

[31] Ioanna Lykourentzou, Angeliki Antoniou, Yannick Naudet, and Steven P Dow. 2016. Personality matters: Balancing for personality types leads to better outcomes for crowd teams. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*. ACM, 260–273.

[32] Jennifer Dodorico McDonald. 2008. Measuring personality constructs: The advantages and disadvantages of self-reports, informant reports and behavioural assessments. *Enquire* 1, 1 (2008), 1–19.

[33] Walter Mischel. 2013. *Personality and assessment*. Psychology Press.

[34] Isabel Briggs Myers, Mary H McCaulley, and Robert Most. 1985. *Manual, a guide to the development and use of the Myers-Briggs type indicator*. consulting psychologists press.

[35] Daniel Olguın Olguın, Peter A Gloor, and Alex Sandy Pentland. 2009. Capturing individual and group behavior with wearable sensors. In *Proceedings of the 2009 aaai spring symposium on human behavior modeling, SSS*, Vol. 9.

[36] Chanda Phelan, Cliff Lampe, and Paul Resnick. 2016. It's creepy, but it doesn't bother me. In *Proceedings of the 2016 CHI conference on human factors in computing systems*. ACM, 5240–5251.

[37] Brent W Roberts. 2009. Back to the future: Personality and assessment and personality development. *Journal of research in personality* 43, 2 (2009), 137–145.

[38] Ivan Robertson and Militza Callinan. 1998. Personality and work behaviour. *European Journal of Work and Organizational Psychology* 7, 3 (1998), 321–340.

[39] Christine Satchell and Paul Dourish. 2009. Beyond the user: use and non-use in HCI. In *Proceedings of the 21st Annual Conference of the Australian Computer-Human Interaction Special Interest Group: Design: Open 24/7*. ACM, 9–16.

[40] Stefan Schneegass, Romina Poguntke, and Tonja Machulla. 2019. Understanding the Impact of Information Representation on Willingness to Share Information. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, 523.

[41] Lee Taber and Steve Whittaker. 2018. Personality depends on the medium: differences in self-perception on Snapchat, Facebook and offline. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, 607.

[42] Jeffrey Warshaw, Tara Matthews, Steve Whittaker, Chris Kau, Mateo Bengualid, and Barton A Smith. 2015. Can an Algorithm Know the Real You?: Understanding People's Reactions to Hyper-personal Analytics Systems. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. ACM, 797–806.

[43] Ziang Xiao, Michelle X Zhou, and Wat-Tat Fu. 2019. Who should be my teammates: Using a conversational agent to understand individuals and help teaming. In *Proceedings of the 24th International Conference on Intelligent User Interfaces*. ACM, 437–447.

[44] Lingling Xu, Cheng Yi, and Yunjie Xu. 2007. Emotional expression online: The impact of task, relationship and personality perception on emoticon usage in instant messenger. *PACIS 2007 Proceedings* (2007), 79.

[45] Penghang Yin, Jiancheng Lyu, Shuai Zhang, Stanley Osher, Yingyong Qi, and Jack Xin. 2019. Understanding straight-through estimator in training activation quantized neural nets. *arXiv preprint arXiv:1903.05662* (2019).