Can an Algorithm Know the "Real You"?: Understanding People's Reactions to Hyper-personal Analytics Systems

Jeffrey Warshaw[1], Tara Matthews[2*], Steve Whittaker[1], Chris Kau[2], Mateo Bengualid[3], Barton A. Smith[2]

[1]University of California at Santa Cruz, CA, USA, {jwarshaw, swhittak}@ucsc.edu

[2]IBM Research - Almaden, CA, USA, taramatthews@gmail.com, {ckau, barton.smith}@us.ibm.com

[3]IBM Argentina, Cordoba, Argentina, benguali@ar.ibm.com

ABSTRACT

Recent research has developed analytics that threaten online self-presentation and privacy by automatically generating profiles of individuals' most personal traits—their personality, values, motivations, and so on. But we know little about people's reactions to personal traits profiles of themselves, or what influences their decisions to share such profiles. We present an early qualitative study of people's reactions to a working *hyper-personal analytics system.* The system lets them see their personality and values profile derived from their own social media text. Our results reveal a paradox. Participants found their personal traits profiles creepily accurate and did not like sharing them in many situations. However, they felt pressured by the social risks of not sharing and showed signs of learned helplessness, leading them to share despite their misgivings. Further, they felt unqualified to significantly modify their profile contents due to a surprising trust in the "expert" algorithm. We explore design implications for hyper-personal analytics systems that consider the needs and preferences of the people being profiled, suggesting ways to enhance the control they feel and the benefits they reap.

Author Keywords

Hyper-personal analytics; privacy; self-presentation; social media; personality; values; user study.

ACM Classification Keywords

H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

INTRODUCTION

A growing body of research is developing analytics to learn all about individuals' *personal traits* [19]—their personality [15,40], values [4], attitudes [12], and so on—to drive *hyper-personalized* technology user experiences. These new hyper-personal analytics threaten both online self-presentation and privacy, as algorithms sift through people's online behavior to uncover implicit personal traits, which individuals may be unaware they are communicating. With these analytics, anyone with access to text you have posted online could generate a profile of your most intimate personal traits. Market researchers and employers have demonstrated a strong interest in personality profiling, long using small-scale personality profiling as a proxy for consumer behavior [26,39], or to improve hiring decisions [3]. Modern analytics systems have made inferring these traits scalable, and marketing and human resources groups now seek to deploy them widely. As such profiles become more common and intimate [15], it is critical to understand how the people being profiled feel about this technology, and how they would want the analytics designed and their outputs shared or kept private.

People are already targets of systems that profile them for ads or personalization based on their online behavior [28,29,37]. Recent work has studied how people feel about targeted behavioral profiles of them created by data aggregators, who compile large amounts of information from various sources to identify an individual's demographics, online behaviors, and interests [29,37]. However, the new class of hyper-personal analytics is very different. Rather than just demographics and interests, they aim to be deep-delving portraits of individuals' personal traits and motivations.

We present a qualitative user study of this new class of hyper-personal analytics system. We explore people's reactions to and preferences for interacting with a working system that assesses their *personality* and *values* from text they have posted in social media, similar to [15]. Prior work studying a precursor system highlighted two main user concerns: others' misinterpretation of the inferred personal traits, and users wanting control to modify their profile and traits [15]. The system we used to explore people's preferences addressed these two concerns: it provided rich definitions of people's traits in paragraph format, which our pilot participants found easy to interpret. It also added functionality to enable people to modify their profile.

CHI 2015, April 18–23, 2015, Seoul, Republic of Korea.

ACM 978-1-4503-3145-6/15/04…$15.00.
http://dx.doi.org/10.1145/2702123.2702274

* Current affiliation: Google Inc.

Our method involved participant use of the study profiling system with their own social media posts during an interview. The profiling system was presented to participants as *voluntary, i.e.,* they had the choice to create and share their profile. We presented participants with exercises evaluating and modifying their profile. Participants also evaluated hypothetical scenarios where they had a choice of whether to share their profile, and a real scenario choosing whether to share the profile on their social media account. This enabled us to address the following research questions:

- How do participants *feel* about using a system that auto-generates a profile of their personality and values, from text they posted in social media?
- How *accurate* is the profile? Does it mirror their real and/or online selves?
- How and to what degree do they *modify* the profile and feel *control* over its contents?
- In what situations do they *share* the profile and to what degree do they feel *control* over sharing?

This paper explores the needs and preferences expressed by people viewing, modifying, and sharing their personal traits profile auto-generated by a working hyper-personal analytics system. Our results reveal a paradox: participants found their personal traits profiles creepily accurate and felt uncomfortable sharing them. However, they felt pressured by the social risks of not sharing and showed signs of learned helplessness, leading them to agree to share their profiles in several scenarios despite their discomfort. Further, they felt greater confidence in the algorithm than in their ability to describe their own traits, rarely modifying what they perceived as the algorithm's "expert" assessment. Even though all study scenarios involved purely voluntary profile creation and sharing, multiple participants spontaneously expressed worries that companies could give their profile to a third party without consent or, in many cases, even profile them without their knowledge. We synthesize these results into design implications for hyper-personal analytics systems that provide greater benefits and control for the people being profiled.

RELATED WORK

We review the two main categories of traits derived by the system in the current study, *personality* and *values*; provide context for our study by reviewing other studies of profiling systems; and ground our study in data privacy findings.

Personal Traits Profiled

Two trait taxonomies that have been studied extensively are the Big-5 personality traits (hereafter, Big-5) and Schwartz's basic human values (hereafter, Values). The Big-5 describes five main personality traits for individual differences in affect, behavior, and cognition: Extraversion, Agreeableness, Conscientiousness, Neuroticism, and Openness-to-Experience [14]. These traits have been studied in multiple domains, showing relations between Big-5 and work performance [3], consumer behavior [11], and relationship success [8].

In contrast to the Big-5, Schwartz's Values describe the beliefs and motivations that guide a person throughout their life. These include four main types of motivations: Self-Transcendence, Openness-to-Change, Conservation, and Self-Enhancement [31]. Schwartz's Values have been validated across 20+ countries [31], and influence a variety of behaviors, including environmental and consumer behavior [26].

Deriving Hyper-personal Traits from Social Media Data

Much prior research has focused on identifying the best predictors for consumer behavior. *Consumer profiling* is an effort to link an individual's buying behaviors to aspects of the individual that might predict their future behavior, such as demographics and interests [39]. With the increase in consumer data available thanks to online shopping and social media, consumer profiling has increasingly incorporated attempts to infer more deeply personal traits based on online behavior through *user modeling*. The last few years have brought several studies that predict an individual's Big-5 personality traits or Values from a variety of social media data sources, including Facebook likes [19], and linguistic analyses of posted text [4,13,15].

Online Self-presentation Management

The proliferation of social networking sites has given rise to a large amount of HCI research on the ways that people manage their self-presentation online [10,22,34]. Although it is difficult to completely control the implicit cues one communicates in-person (with tone, facial expressions, body language, and so on), prior research argues that the ability to self-curate social media profiles gives people control over the implicit cues others can interpret from their online presence [34], usually by emphasizing positive traits and downplaying negative traits [22]. Hyper-personal analytics systems that automatically infer implicit traits may therefore be seen by people as a threat to attempts to actively manage their self-presentation online.

People's Reactions to Profiling

Many studies have explored people's understanding of and reactions to their data being used to profile them [28,29,37], but these were based on more explicit characteristics: demographics, behaviors, and interests; not personal traits. People show an overall discomfort with being profiled [29], but they also underestimate the depth of inferences that can be drawn from online behavior [28,29]. Only one study we are aware of has explored people's reactions to a hyper-personal analytics system that assesses personal traits, a survey by Gou et al. [15] during which survey respondents interacted with a system that presented a visualization of their automatically-inferred Big-5 and Values. Some respondents expressed a willingness to share their inferred traits in a work setting but were concerned about coworkers misinterpreting the visualization, asking for the ability to correct perceived inaccuracies in the results.

We provide rich, qualitative data about perceived accuracy and acceptability of a hyper-personal analytics system that profiles personality and values derived from social media posts. We also contribute observations of how people modify and share their profile.

Privacy Calculus: Sharing Personal Information

Privacy is the process by which people manage boundaries and disclosure across contexts [25]. Importantly, privacy *attitudes* can differ radically from actual privacy *behavior* [1]. People who claim to have strong general data privacy concerns nonetheless share their data when they see benefits for doing so [6]. This decision-making process is driven by a *privacy calculus*, in which people weigh benefits, risks, and perceived control when considering sharing their data [21]. Because of this, we designed our study so that participants were responding to specific profile sharing scenarios.

People perceive high *benefits* to sharing situations that provide them with new opportunities to enhance their current circumstances. Some benefits to sharing personal information include facilitating socialization [2]; financial benefits, e.g., discounts [9]; and increased convenience, e.g., keeping a credit card on file with an online merchant.

Privacy *risk* factors include concerns about security, third-party sharing, and the potential for social criticism. The possibility of unintentional data leakage is a large security concern, but people are also highly concerned about their data being *purposely* shared with third parties without their consent [30]. Third-party sharing is associated with risks in several domains: in commerce, receiving spam; in work settings, negative ramifications for career or work socialization [38]; in social sharing, humiliation [7]. There are also potential risks attached to *not* sharing: pressure to participate in online networks from peers [16] or work colleagues [38] often leads people to share more than they would like. Perceptions of risk can be affected by system design, for example, calling attention to the specific data at risk when users are making a sharing decision improves risk awareness [17].

Several privacy theories consider perceived *control* over one's data as affecting sharing decisions [21]. People who experience privacy leakages perceive greater risks to sharing and hold their personal information more closely in the future [7]. Such leakages are often accompanied by a sense of "creepiness": an apprehensive, unsettling emotional response to an ambiguous threat to one's sense of control [23,37].

Because of the complexity of privacy decisions, we presented participants with scenarios where they could experience a variety of configurations of benefits, risks, control, and sharing norms, to explore how each of these factors contribute to their willingness to share personal traits profiles.

METHOD

We asked 18 participants to use a hyper-personal analytics system that generated a profile of (a) their Big-5 personality traits [14] and (b) Schwartz's basic human values [31]. Participants logged into their Twitter or Facebook account using this study system and watched their personal traits profile auto-generate from the text they had previously posted. They were first asked to evaluate their profile for accuracy and next given scenarios—both hypothetical and real—in which they were asked to consider modifying and sharing their profile. We conducted semi-structured interviews as participants used the system and analyzed for themes that emerged from their interactions with the system and the scenarios.

Participants

We interviewed 18 participants (9 female, 9 male), ranging in age from 19-59. Participants needed to be either Twitter or Facebook users with enough content for linguistic analysis. Our participants all had >195 tweets or >80 Facebook posts. All participants were recruited from a single global enterprise business at locations in the San Francisco Bay Area. They represent a variety of job roles, including managers, interns, research staff, sales, communications, and administrative assistants.

System

To allow participants to voluntarily create their profile, we modified an existing experimental system, *KnowMe* [15]. It allows a person to login to a Twitter or Facebook account and see their automatically inferred personality traits based on the text they have posted there. The system correlated the relative frequency of categories of words used by the individual with associated Big-5 and Values traits, using a dictionary based on LIWC [35]. The links between these word categories and traits have been previously described in [4,40]. For Twitter accounts, the system pulled up to 200 of the most recent non-retweet tweets; for Facebook, up to 100 of the most recent status updates, wall posts, or comments. Prior work indicates this is enough text for a representative language sample [15].

Our system presented people with a *profile paragraph* of automatically-generated text based on their derived traits (see Figure 1). We arrived at this design following a user-centered design process involving multiple rounds of pilot testing and design iteration to transition from the visualization in [15] to the profile paragraph. In all, 14 pilot participants used *KnowMe* and provided feedback until we arrived at the final design. Initial pilot participants found the visualization difficult to interpret, especially given the complexity of the traits information shown; they felt they needed more explanation on how to interpret the multiple traits together. Pilot participants were most interested in their most extreme Big-5 traits (in relation to the population median), their highest Values, and how their traits combined to create their personality, so we focused the profile paragraph on those. The paragraph therefore

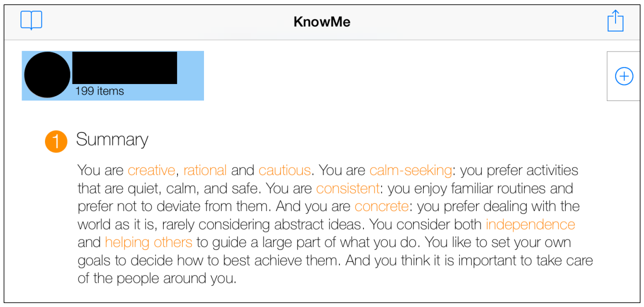

Figure 1. User interface of the study system, showing a derived personality and values profile (P15).

contained 3 combined elements: 3 *cross-trait* Big-5 personality terms; 3 Big-5 personality *facet* (subtype) terms and descriptive sentences; and 2 basic *values* terms and descriptive sentences. The cross-trait terms and facets together captured the individual's Big-5 traits. The generated terms are precise; the Big-5 traits are commonly used in psychological and clinical practice [20]. They also include a balance of positive and negative terms (e.g., *cowardly*, *explosive*, *unscrupulous*). We describe how profile terms were generated in the next three sub-sections.

Cross-trait Terms

Pilot participants wanted to better understand how their Big-5 traits affected each other, leading us to add *cross-trait* terms to the final profile paragraph. This is also a more accurate way of describing an individual's personality than focusing on isolated traits, because overall personality involves the combination of multiple traits [14,18].

To create cross-trait terms, we surveyed the Big-5 literature compiling a list of adjectives that have been validated as each describing a specific combination of two Big-5 personality traits (e.g., high agreeableness and high extraversion may be described as *sociable*, high agreeableness and high neuroticism as *sentimental*) [14,18]. For combinations of traits that were uncommonly described in prior literature (e.g., high neuroticism with low openness-to-experience), we created terms and validated them using Mechanical Turk in a procedure similar to [24]. Several of the Turk-validated terms tend to be viewed negatively in Western culture (e.g., *insensitive*, *high-strung*), but we included them to prevent biasing participants by only showing positive trait terms. The adjective list we generated enabled us to engage with our research questions, so it is useful but not exhaustive or complete.

Following recommendations of our pilot participants, the profile paragraph first provided 3 terms describing a person's most extreme Big-5 traits using a moving window: we showed the term for the first and second most extreme 2 traits of that user, then the second and third most extreme 2 traits, followed by the third and fourth most extreme 2 traits. Although some Big-5 traits can be interpreted as having only one good extreme (e.g., high agreeableness), we presented the most extreme traits derived by the system, rather than the most positive traits. Terms that are viewed negatively in Western culture were preceded by a qualifier (e.g., *a bit*, *somewhat*) to provide the user with an honest but slightly hedged description. For example in Figure 1, high Openness and low Neuroticism generated the term *creative*; low Neuroticism and high Conscientiousness generated the term *rational*; finally, high Conscientiousness and low Extraversion generated the term *cautious*.

Facets

Several Big-5 models subdivide traits into facets, subtypes that more specifically describe a predisposition towards an attitude or behavior. We used the IPIP facets [14], which split each Big-5 trait into six facets (e.g., agreeableness: trust, cooperation, altruism, morality, modesty, sympathy). We generated terms as well as descriptive sentences for the high and low ends of each facet based on the literature [14].

In the profile paragraph, a user's three most extreme facets were assigned a term and descriptive sentence based on being higher or lower than the population median. For example in Figure 1, the participant's low *excitement-seeking* facet generated the sentence beginning "You are *calm-seeking*...".

Values

Based on [31], we generated terms and descriptive sentences for several values that lay within each values cluster (e.g., the *self-enhancement* cluster contains *gaining social status, ambition, achieving success, high achievement*). For the profile, the system showed the individual's strongest values terms within a descriptive sentence. For example in Figure 1, the participant was strong on Openness-to-Change and Self-Transcendence values, generating the text beginning "You consider both *independence* and *helping others*...".

Interviews

We conducted semi-structured interviews in which 13 out of 18 participants did the following:

1. Used the system to create a *personal traits profile* (on an iPad provided by the interviewer).
2. Examined and discussed the accuracy of, and their reactions to, their profile.
    a. Completed a survey rating the accuracy and valence of each trait in their profile (*survey on profile traits*).
3. For each of 3 to 8 hypothetical and real sharing scenarios:
    a. Modified their profile for that scenario (*profile modification activity*).
    b. Made a decision on whether or not to share (*profile sharing scenarios*).

A control group of 5 participants followed the procedure above minus the ability to modify their profile. The goal of this control group was to compare participants' perceived control in both conditions. Because few participants

Scenario	Characteristics	Proportion Who Shared
Event Recommender	Non-work, non-reciprocal	.93 (N=14)
Brainstorming Group	Work, non-reciprocal, reported in aggregate	.78 (N=9)
Speed Mentoring	Work, reciprocal	.75 (N=16)
Hobby Group	Non-work, reciprocal	.73 (N=11)
Team Formation	Work, reciprocal	.67 (N=15)
Online Shopping Discount	Non-work, non-reciprocal, shared with third-party	.64 (N=14)
Job Application	Work, non-reciprocal	.59 (N=17)
Actual Account	Shared on user's real social media account	.56 (N=16)

Table 1. Participants were generally willing to share in multiple scenarios.

changed items in their profile, we included more people in the experimental group to explore this in more detail.

The first author conducted and coded all interviews. Interviews were face-to-face, lasted 30-80 minutes, were audio recorded, and detailed notes were taken. After the interviews, we reviewed the audio to enhance the notes. We analyzed the notes using open coding (22 codes total), and then iteratively analyzed the concepts and categories from our initial coding for themes, which we report.

In the following three sub-sections, we detail the survey on profile accuracy and valence (2a), the profile modification activity (3a), and the profile sharing scenarios (3b).

Survey on Profile Traits

Participants rated each trait shown in their profile for three questions, "*How accurately does '[trait]' describe you?*", "*How positive (good, praising) is the term '[trait]'?*", and "*How negative (bad, critical) is the term '[trait]'?*". Accuracy was rated from (1) "Not at all accurate" to (4) "Completely accurate". Positivity/negativity were rated from (1) "Not at all [positive/negative]" to (4) "Strongly [positive/negative]".

Profile Modification Activity

Prior findings show that people want to modify traits in their profiles [15]. Pilot participants found it difficult to consider the implications of freeform changes, so we built three types of modifications into the study system: *hiding, changing words,* and *adding qualifiers*. So as not to overwhelm participants with choices, we implemented these three modifiers on subsets of the profile where we believed they would be most understandable and useful. *Hiding* removed a trait from view entirely and could be performed on any trait in the profile. *Changing words* was enabled for the cross-trait terms and values elements, so that users could see and choose previously validated alternatives to the terms and descriptions. *Adding qualifiers* was enabled for the Big-5 facet terms. Participants could add understating (e.g., somewhat, sometimes) or exaggerating (e.g., very, quite) qualifiers to any facet term.

Profile Sharing Scenarios

Up to 8 scenarios were presented to participants, each representing a specific situation in which they were asked whether they would voluntarily share their automatically-inferred personal traits profile with the party requesting it. These scenarios systematically explored how different situational contexts affected sharing decisions: audience, reciprocity, third party disclosure, monetary rewards, and recommendations (see Table 1, column 2). We consulted with industry HR professionals to ensure the wording for the *job application* and *team formation* scenarios were true-to-life. Based on prior research, we expected an audience effect, with people showing lower willingness to share with a team leader or hiring manager due to the high risk involved [38]. Prior research on peer pressure and unraveling [16], predicts higher willingness to share in scenarios when the requesting party shares their profile in return.

We present the *job application* scenario in full, then outline the other scenarios. For each, participants were given the same verbal instruction to talk through their thoughts as they decided whether or not to share their profile:

> *Job application: You are interested in applying to a job posting at an organization that you respect. In order for your application to be considered, you are required to submit a [personality profile]. The job posting mentions that it gives them a better understanding of how you might fit in with the existing organizational culture. It also says that the [personality profile] will be an important part of the evaluation process for your application, but it will not be the sole basis for the hiring decision. If you are hired, your [personality profile] will remain in your employment file with your other application materials.*

The *job application* scenario probes whether the participant will share with a *work* audience to be considered for a job. The organization requesting the profile does not provide personality profiles of current employees so it is classified as *non-reciprocal*.

Three other work scenarios were presented (see Table 1): (1) *Team formation*: sharing with a new work team leader prior to a first group meeting. Participants were shown the leader's profile before deciding (reciprocal). (2) *Brainstorming group*: organizers of a work event to brainstorm ways to improve client service asked for the participant's profile to aggregate with other attendees', in order to understand who takes part in such events (non-reciprocal). (3) *Speed mentoring*: sharing to facilitate meetings with potential career mentors. Participants would receive each mentor's profile after deciding whether to share, but before the meeting (reciprocal).

Three non-work hypothetical scenarios were presented: (1) *Event recommender*: sharing their profile with a company to receive activity recommendations after moving to a new city, based on activities that people similar to them had enjoyed (non-reciprocal). (2) *Hobby group*: the organizer of a hobby group asked participants to submit profiles, promising to share the group's average personality profile for the participant to assess their fit with the group (reciprocal). (3) *Online shopping discount*: sharing their profile with their favorite online store for a 50% discount at a favorite online store, with the provision that their profile may be shared with or sold to the store's partners for marketing purposes (non-reciprocal).

Finally, in the *actual account* scenario, participants were invited to post their profile to the social media account from which the profile was derived. This allowed us to compare participant decisions to share in hypothetical and real scenarios.

RESULTS

Participants found their auto-generated personality profile 'creepily' accurate and expressed discomfort with sharing it in many situations. Nevertheless many said they would share their profile despite strong reservations, and did not modify its contents. Their rationale for these somewhat contradictory reactions revealed several themes. Though participants discussed a series of *risks* and *benefits* to modifying and sharing their profile, the decision was considerably more complex than a simple weighing of the two. Participants felt unqualified to significantly modify their profile contents due to a surprising trust in the "expert" algorithm. They also felt social pressure to share to prove they had nothing bad to hide, coupled with a learned helplessness—a belief that they could not keep their profile private even if they wanted to. Despite the voluntary nature of all presented scenarios, several participants spontaneously expressed worries about potential involuntary deployments of hyper-personal analytics systems. Below, we distinguish feedback based on the presented scenarios from that based on participant speculation about third-party appropriation.

Reactions to System-Inferred Personal Traits Profiles

Accuracy

Participants described the system-generated profiles as surprisingly accurate. 17/18 stated their profile was an accurate representation of their general characteristics as an individual. They discussed the traits in relation to both their online personas and perceptions of their actual selves:

> *"I'd say this is who I am on Twitter for sure. I'm maybe more emotional in my personal life, but things like the [trait 'respecting authority'] would transition over." (P16)*

Even participants who used Twitter primarily for professional reasons, described the profile as characterizing both their online and their offline personality and values:

> *"I'm surprised, actually, that it's really accurate. This is the kind of discussion you usually have with friends or a spouse over a couple of beers... I was thinking, 'my Twitter is a professional account with 1200 followers.' You tend to be a little more professional [and] craft things a little more carefully. It kind of came up with an analysis that's really accurate." (P15)*

In addition to describing the profile's general accuracy, we asked participants to rate each trait in the profile for accuracy, positivity, and negativity. The rating revealed high accuracy overall for cross-traits (M = 3.04, SD = .84), facets (M = 3.44, SD = .73), and values (M = 3.50, SD = .89). Accuracy ratings were moderately correlated with positivity ratings for personality terms ($.30 < r\text{s} < .47$) and highly correlated for values terms ($.72 < r\text{s} < .80$), echoing psychology research that people view their own traits positively, e.g., [36].

Creepiness

Creepiness was a major theme in participants' reactions to the system's accuracy. Nearly all participants were surprised at the system's accuracy at inferring personal traits. However half (9) explicitly said they felt uncomfortable that it penetrated their carefully crafted online presence to derive fundamental traits like personality and values. Participants described feeling creepiness over two system aspects: the *personal and implicit inferences* it made, and although the study scenarios gave them the power to not share, the worry that similar systems could be used on their data *without their consent*.

Personal and implicit inferences. Participants expressed a visceral discomfort with their profile revealing personal traits they thought were absent from their online presence. For example, P8 was particularly taken aback by the *values* section of the profile, which was "eerie" in its accuracy given what he felt was a small number of data points (196 tweets):

> *"I don't know how it would derive that from the limited number of tweets that I made... I always thought of that as something that I don't let people on to know very much... I guess I'm a little shocked that it works so well. It's not just like a horoscope or something like that." (P8)*

Participants also described this system as being more invasive than others. These other systems were seen as drawing more *surface level* inferences about them, relating to their consumer behavior rather than personality or values:

> *"It's kind of creepy... I know they do it at the grocery store with your card, Facebook does it, web pages do it... whereas this, I feel like I'm being tracked." (P18)*

Profiling without consent. Even though the current scenarios were framed specifically to allow participants to decline sharing, they were aware and highly uncomfortable that companies are already able to profile them without their consent or notification. After discussing the scenarios, P10 described unease at unknown third parties using the system to summarize his character in a short blurb:

> *"Allowing other people to go, 'oh yeah, that's who you are.' That's creepy." (P10)*

Participants felt strongly that their consent to infer these hyper-personal traits should be required, regardless of the privacy settings on their social media posts:

"It has to be opt-in. Even if [my Twitter account] is public, I am controlling what I'm putting out. Even if there's an app [letting people look each other up], I wouldn't allow something to be done outside my knowledge. It has to be opt-in, has to be done in that context." (P15)

(Non-)Modification of Personal Traits Profiles

After logging in and seeing the study system generate their personal traits profile, participants engaged with a series of scenarios (see Table 1), during which they were offered the chance to modify any of the eight items in the profile before deciding whether or not to share it. To our surprise, however, participants made *very few or no changes at all* to their profile. This was despite all participants understanding how to modify their profile. Of the 13 participants who were given the option to alter aspects of their profile, 6 did not change it for any scenario, 4 changed one item, 2 changed two items, and 1 changed four items. The changes included hiding traits (6 changes), adding understatement qualifiers such as 'somewhat' or 'slightly' (4 changes), and changing words (2 changes). All participants who modified their profile submitted the same changes across scenarios, noting that it was a good representation of their personality and values, and they saw no need for multiple versions.

The surprising accuracy of their profiles seemed to have instilled participants with faith in the system. This faith had an unanticipated effect: participants were reluctant to override the default word choices the system provided. The most common explanation participants gave for not changing their profile was *insecurity that they could craft a better profile than the system.* This feeling of insecurity focused on *changing words;* participants felt more confident when *hiding* or *qualifying* terms.

When participants were ambivalent about how to modify the words in their profile, they consistently trusted the algorithm over their own judgment:

"To me, I don't know who's looking at it. I don't know what they're going to be valuing... If it's a government job, maybe I don't want to include 'laid back'. At the end of the day, that's who the system says I am. There's obviously been some type of process where they think this is useful." (P6)

"Each word means something to different people... It's almost impossible to know how others would perceive your profile. I don't think it's ever possible for me to write in a few sentences something that describes me 100%... And so I'm not sure how to change something to make it more accurate." (P2)

Sharing System-Inferred Personal Traits Profiles

Before and after modifying their profile for each scenario in Table 1, participants were asked to decide whether or not they would share it. Overall, participants did not feel completely comfortable about, or in full control of, their decision to share their profile. Nevertheless they very frequently stated they would share it anyway, whether they modified it or not. Across all hypothetical scenarios, participants shared a mean of 70% of the time. When they were sharing on their *actual* Facebook or Twitter account, 56% of participants shared. These high levels of sharing despite some discomfort and relatively little modification appear contradictory.

While one group of 13 participants could modify their profiles, 5 participants were not offered this ability. There was no noticeable difference between the two groups in sharing decisions, but participants who could change their profile felt more comfortable sharing the modified profile over the original, even though these changes were extremely minor.

Prior work describes people's decisions to share private information as a complex *privacy calculus*, which weighs risks and benefits [21]. An individual's privacy calculus can result in sharing decisions that seem to contradict their stated privacy concerns and preferences [1,7]. To fit this framework, we present our results on participants' sharing decisions in terms of *risks* and *benefits.*

Before presenting risks and benefits, we note that some participants felt the scale was tipped toward the risks but they were helpless to opt-out, even though our scenarios explicitly provided them with this possibility. This is a form of *learned helplessness*, in which people feel they have no power to avert an avoidable negative outcome [32]:

"Knowing other companies mine [social media data] is creepy, but it's unstoppable." (P10)

"It's a little eerie, our tweets are public domain unless your [account is] private. I don't see it would be anything that you could stop or prevent. If I could block people from using it, I would." (P8)

Risks

Three main themes of risks for voluntary sharing emerged: concerns about prejudging based on profile contents, the potential for disclosure to an unknown third-party, and ascribing negative attributes to non-sharers.

Concern about prejudging. The most common concern was being *prejudged*. Participants worried that someone reading their profile would prematurely assess their character before meeting them. A subset of these participants also expressed concerns that profiles might be out of date if they had not posted enough content recently:

"I want them to understand my personality when I join and grow rather than them going with something historical." (P15)

Prejudging was also something participants worried about doing to others in reciprocal scenarios (e.g., team formation, speed mentoring). This participant mentioned a risk of *prejudging others*, explaining others' behavior by invoking their profile:

"I would be worried I would look for those traits. When I would talk to you: 'Is that because he's guarded, or because he doesn't like me?'" (P7)

Third-party disclosure. Participants frequently and spontaneously shared concerns about their profile being shared more widely than the scenario described. They discussed third-party marketers in the non-work scenarios, or managers and peers in the work scenarios. The *online shopping discount* and *job application* scenarios were seen as particularly high risk and had the lowest sharing rates of the hypothetical scenarios (see Table 1). The concern over third-party disclosure confirms results from prior studies of other types of personal information sharing decisions [30].

Ascribing negative attributes to non-sharers. Many participants felt compelled to share their profile to avoid others assuming they were hiding something negative, and being excluded from benefits that sharers would accumulate. In the law literature, this is termed the "unraveling effect" [27]: non-sharing is interpreted as *hiding terrible information*, pressuring non-sharers to share against their wishes. Participants expressed this social pressure both directed at themselves and at others:

> *"[It's] forcing, almost. If they know you decline, that's more of a red flag to them." (P16)*

> *"[If someone hid], I would want to know, what horrible thing was it? You immediately go to the negative." (P18)*

Benefits

Participants ascribed higher benefits to sharing when the scenario involved them getting something in return. This is often described in the literature as exchanging personal data as a *digital currency* [9]. Some participants said they *wanted* to exchange their traits profile for reciprocal sharing, recommendations, and a "trusted source" (the system) to vouch for implicit aspects of their character.

Reciprocal sharing. Although three of our scenarios involved reciprocal sharing, in only one was reciprocity seen as a benefit that would motivate participants to share. The *team formation* scenario enabled participants to make assessments about how well they might fit in a new work group and interact with other members. On the other hand, participants felt that personality profiles were not as useful in other reciprocal sharing scenarios (*speed mentoring* and *hobby group*); instead, other qualities about mentors and fellow hobbyists were more important, like their expertise. In *team formation*, enough value was perceived, so the reciprocity of the sharing became important:

> *"It increased [my willingness to share]. They're asking you to do something that they're already doing." (P6)*

The implication is that reciprocity can motivate people to share, but receiving others' personality information must offer significant value.

Personality-based recommendations. Participants were overwhelmingly positive about sharing their profile to get activity or event recommendations based on people who were like them (e.g., *event recommender* scenario). Participants explained that they believed personality-matched recommendations would be better than seeing just unfiltered recommendations:

> *"I think this is awesome. If you look at yelp, someone gives [a business] 5 stars, you don't know who that person is or their story." (P6)*

Trusted source vouching for implicit aspects of one's character. Nearly half of participants (8/18, primarily those in service or sales job roles) expressed a strong desire to share their personal traits profile in work settings. For these participants, the study system represented a *trusted source*, since it inferred traits implicitly rather than via self-report, allowing others to verify traits that were hard to demonstrate prior to meeting. This was discussed most frequently in the *job application* scenario:

> *"I wish job applications had this kind of submission. Someone like me who has a lot of between-the-lines interpersonal skills that are hard to build into a resume to show you have that depth. To me, this is what gives people depth and dimension to not only find fit but also give potential employers an idea about what you have as a person... I would much prefer using this data than softer data points about analyzing the way they're talking to me, clothes they're wearing, all the social cues. Nice to go into it, "you've been certified THIS." (P13)*

DISCUSSION

Although there has been much study of user-curated online profiles, we present people's reactions to using a system that automatically generates a hyper-personal profile of their traits and values [15]. We found that people are highly sensitive to the social risks of not sharing and showed signs of learned helplessness, often sharing what they felt were deeply personal, creepily accurate profiles despite strong discomfort. People were more comfortable with sharing their profile when they were able to modify it—even though they did not feel empowered to make many changes. People were acutely aware of the value their profile holds to others, often sharing to get some benefit by treating their personal data as a digital currency. Experiencing the system and scenarios prompted several participants to express discomfort that companies could profile them without their permission or knowledge. This, combined with the high proportion of sharing in our *voluntary* scenarios, provides additional support for online privacy theories that emphasize that honest disclosure about user data transactions stimulates user comfort with sharing [21].

One of the surprising results from our study was that most participants put more faith in the system than was likely warranted given its still-experimental nature and assessment of very personal traits. They perceived high accuracy even when the system presented participants' negative traits. Again we saw a lack of perceived control. In ambiguous situations, participants trusted the system's word choices *more than their own preferences*. It is fascinating and alarming the degree to which participants trusted the study system to provide a 'character reference'. That an algorithm might be viewed as an *expert on human character,* to officially vouch for a person, puts the computer in a position previously held by people. Granting the algorithm this level of "humanity" simultaneously reduces our humanity by supplanting people as the sole judges of character. This result raises the ethical question: should an algorithm judge character?

Our participants felt a strong sense of "creepiness" when extrapolating from the study that companies could profile them without their permission using their social media text. In daily life, our personality traits and personal values are implicitly communicated through our behaviors and interactions with others. But such intimate information can be difficult to control in person, with people often feeling better able to curate their self-presentation online [34]. This sense of greater control over online self-presentation may have been undermined when participants saw exactly what the algorithm, and therefore companies that may deploy it, could infer about them. Furthermore, inferring personal traits is generally the purview of other humans. These reasonable concerns could have heightened participant's feelings of creepiness about the study system.

Design Implications

Systems that allow transparency and control over auto-generated profiles are fairly rare [29]. However, people are aware of and uncomfortable with companies deploying profiling systems. This places a requirement on designers to consider users' reactions to their systems, whether being profiled is voluntary or involuntary. Our participants offered insights into future designs which might be instantiated by *building responsible system defaults* [33]. We also provide user-driven guidance on deployment features that would provide people with a greater confidence and comfort with profiling: control over the *content* of derived insights about them, and control over *sharing*. Although these general topics have been studied in prior online privacy work [21,29], our findings indicate that hyper-personal systems are viewed as qualitatively different from traditional consumer profiling systems. Our guidelines therefore address the specific challenges these new systems raise.

Trust in Analytics and the Importance of Defaults

Our results show the importance of default profile wording, and its role in building trust in an analytics system users interact with directly. Participants' initial encounter with a profile they felt accurately described them seemed to instill immediate trust. Including negative traits increased trust further: participants who evaluated a derived trait as both accurate and negative thought it boosted the accuracy of the system beyond a positive item. This finding about people's trust in the system's initial profile extends prior research on the strong effect of system defaults on user behavior [33], showing this effect holds, surprisingly, when people are evaluating descriptions of *their own personality and values*—a topic on which we would expect the person profiled to believe themselves to be the ultimate expert. Thus, decisions that system builders make regarding the words displayed about traits can significantly impact user experience.

Instilling Control to Modify Derived Profile Contents

Our participants experienced a lack of control over the contents of their personal traits, which reduced their willingness to correct inaccuracies in their profiles. It is therefore critical that designers foster people's ability and confidence to modify their profiles. This is especially important since others might make decisions about a profiled person based on the profile a person shares. First, the fallibility of the algorithm and the potential need for modifications should be communicated. For example, confidence levels for certain terms could be included. This might be suggested by simple visuals such as graying low confidence words or by presenting multiple alternative terms allowing users to choose between them. Second, people's concerns about *hiding terms* and *changing words* were social in nature—they were worried making these modifications could insinuate something negative about their personality—so a social counter-argument might help. For example, the ability to hide traits could be accompanied by statistics about how many others had decided to hide each trait, or by showing how many edits a user typically made to their profile. This type of *social proof* encouragement has been shown to motivate action [5].

Control over Derived Profile Sharing

In line with prior work [21], participant feelings about profile sharing followed a reliable calculus that was *situation-specific*: high benefit scenarios and high perceived control over sharing promoted sharing, but high risk scenarios decreased it. Designers can foster the greatest sense of control over sharing profiles by offering people strong benefits for sharing, and by providing people explicit power to decide whether or not to share their profile.

Our study indicates which use cases designers should target to offer the most benefit to the people profiled: providing *personality-based recommendations*, offering *discounts, and* supporting *reciprocal sharing of personal traits profiles*. Recommendations that, for example, gave insights into the values or personality of an online reviewer, were seen as particularly positive. These represented a new type of filter for products or events that added to other filters they currently use (e.g., social recommendations). Discounts were seen as providing high economic value for sharing, but were still shared less frequently due to the high risk they represented. Reciprocity was seen as a positive when the user received the other party's profile prior to sharing. However the risk of *unraveling* was very salient, and participants felt pressure to share to avoid assumptions that they had something negative to hide. By providing more information about *contexts of use,* designers could more clearly outline these benefits and potential risks, allowing people to make more informed decisions about profile sharing. Our results indicate that people associated different values to the benefits and risks and should be empowered to choose whether or not their profile is shared. For example, some would exchange it with companies for better recommendations, but others were strongly opposed.

CONCLUSION

New systems that enable companies to generate new types of hyper-personal profiles raise important issues. User reactions to such a system identified significant concerns: the *creepiness* of being accurately profiled by an automatic system, and the desire for *control* over profile contents and privacy. These concerns are complicated by a tendency for people to trust the supposed *impartiality* of an algorithm, which fostered unwillingness to correct profile errors. Future research needs to further investigate user reactions to such systems as they evolve. We also need to develop and evaluate new designs that provide greater user control over hyper-personal profiles, supporting editing and informed sharing, as this type of profiling is deployed more widely.

REFERENCES

1. Acquisti, A., Grossklags, J. Privacy attitudes and privacy behavior. In *Economics of Info. Security*. '04, 165–178.
2. Barkhuus, L., Tashiro, J. Student Socialization in the Age of Facebook. *Proc. of CHI*, ('10), 133–142.
3. Barrick, M.R., Mount, M.K. The big five personality dimensions and job performance: A meta-analysis. *Personnel Psychology 44*, ('91), 26.
4. Chen, J., et al. Understanding individuals' personal values from social media word use. *Proc. of CSCW*, ('14), 405–414.
5. Cialdini, R.B. *Influence: The Psychology of Persuasion, Revised Edition.* Harper Business, NY, '06.
6. Connelly, K., Khalil, A., Liu, Y. Do I do what I say?: Observed versus stated privacy preferences. In *Human Computer Interaction - INTERACT*. Springer, '07, 620–623.
7. Debatin, B., Lovejoy, J.P., Horn, A.-K., Hughes, B.N. Facebook and online privacy: Attitudes, behaviors, and unintended consequences. *J. of CMC 15*, ('09), 83–108.
8. Donnellan, M.B., Conger, R.D., Bryant, C.M. The big five and enduring marriages. *J. of Res. in Personality 38*, '04, 481.
9. Egelman, S., Felt, A., Wagner, D. Choice architecture and smartphone privacy: There's a price for that. In *The Economics of Info. Security and Privacy*. '13, 211–236.
10. Ellison, N., Heino, R., Gibbs, J. Managing impressions online: Self-presentation processes in the online dating environment. *J. of CMC 11*, ('06), 415–441.
11. Fraj, E., Martinez, E. Influence of personality on ecological consumer behaviour. *J. of Consumer Beh. 5*, ('06), 167–181.
12. Gao, H., et al. Modeling user attitude toward controversial topics in online social media. *Proc. of ICWSM*, ('14).
13. Golbeck, J., et al. Predicting personality from Twitter. *Proc. of PASSAT & SOCIALCOM*, ('11), 149–156.
14. Goldberg, L.R. A broad-bandwidth, public-domain, personality inventory measuring the lower-level facets of several Five-Factor models. In *Pers. Psych. in Europe*. '99, 7.
15. Gou, L., Zhou, M.X., Yang, H. KnowMe and ShareMe: Understanding automatically discovered personality traits from social media and user sharing preferences. *Proc. of CHI*, ('14), 955–964.
16. Gross, R., Acquisti, A. Information revelation and privacy in online social networks. *Proc. of WPSE*, ('05), 71–80.
17. Harbach, M., Hettig, M., Weber, S., Smith, M. Using personal examples to improve risk communication for security & privacy decisions. *Proc. of CHI*, ('14), 2647.
18. Hofstee, W.K., De Raad, B., Goldberg, L.R. Integration of the big five and circumplex approaches to trait structure. *J. of Personality and Social Psychology 63*, ('92), 146.
19. Kosinski, M., Stillwell, D., Graepel, T. Private traits and attributes are predictable from digital records of human behavior. *PNAS 110*, ('13), 5802–5805.
20. Kotov, R., Gamez, W., Schmidt, F., Watson, D. Linking "big" personality traits to anxiety, depressive, and substance use disorders: A meta-analysis. *Psych. Bulletin 136*, '10, 768.
21. Li, Y. Theories in online information privacy research: A critical review and an integrated framework. *Decision Support Systems 54*, ('12), 471–481.
22. Marwick, A.E., Boyd, D. I tweet honestly, I tweet passionately: Twitter users, context collapse, and the imagined audience. *New Media & Society 13*, ('11), 114–133.
23. McAndrew, F.T., Koehnke, S.S. Creepiness. *Poster presented at SPSP*, ('13).
24. Mohammad, S.M., Turney, P.D. Emotions evoked by common words and phrases: Using mechanical turk to create an emotion lexicon. *Proc. of NAACL-HCT*, ('10), 26–34.
25. Palen, L., Dourish, P. Unpacking "Privacy" for a Networked World. *Proc. of CHI*, ('03), 129–136.
26. Pepper, M., Jackson, T., Uzzell, D. An examination of the values that motivate socially conscious and frugal consumer behaviours. *Intl. J. of Consumer Studies 33*, ('09), 126–136.
27. Peppet, S.R. Unraveling privacy: The personal prospectus and the threat of a full-disclosure future. *Nw. UL Rev. 105*, ('11), 1153.
28. Rader, E. Awareness of behavioral tracking and information privacy concern in Facebook and Google. *Proc. SOUPS*, '14.
29. Rao, A., Schaub, F., Sadeh, N. *What do they know about me? Contents and concerns of online behavioral profiles.* CMU-CyLab-14-011, '14.
30. Schwaig, K.S., et al. A model of consumers' perceptions of the invasion of information privacy. *Info. & Mgmt. 50*, '13, 1.
31. Schwartz, S.H. Universals in the content and structure of values: Theoretical advances and empirical tests in 20 countries. *Advances in Experimental Social Psych. 25*, ('92).
32. Seligman, M.E.P. *Helplessness: On Depression, Development, and Death.* W.H. Freeman & Co., NY, '92.
33. Shah, R.C., Sandvig, C. Software defaults as de facto regulation: The case of the wireless internet. *Information, Communication & Society 11*, ('08), 25–46.
34. Strano, M. User descriptions and interpretations of self-presentation through Facebook profile images. *Cyberpsychology: J. of Psychosocial Research on Cyberspace 2*, ('02).
35. Tausczik, Y.R., Pennebaker, J.W. The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods. *J. of Language and Social Psych. 29*, ('10), 24–54.
36. Taylor, S.E., Brown, J.D. Illusion and well-being: A social psychological perspective on mental health. *Psychological Bulletin 103*, ('88), 193–210.
37. Ur, B., et al. Smart, useful, scary, creepy: Perceptions of online behavioral advertising. *Proc. of SOUPS*, ('12), 4.
38. Wang, Y., Kobsa, A. Privacy in online social networking at workplace. *Proc. of CSE*, ('09), 975–978.
39. Weber, K., Roehl, W.S. Profiling people searching for and purchasing travel products on the world wide web. *J. of Travel Research 37*, ('99), 291–298.
40. Yarkoni, T. Personality in 100,000 Words: A large-scale analysis of personality and word use among bloggers. *Journal of Research in Personality 44*, ('10), 363–373.