“Crystal Is Creepy, but Cool”: Mapping Folk Theories and Responses to Automated Personality Recognition Algorithms

Tony Liao and Olivia Tyson

Abstract
This article examines Crystal Knows, a company that generates automated personality profiles through an algorithm and sells access to their database. These algorithms are the result of a long line of research into computational and predictive algorithms that track social media practices and uses them to infer individual characteristics and make psychometric assessments. Although it is now computationally possible, these algorithms are not widely known or understood by the general public. Little is known about how people would respond to them, particularly when they do not even know their online activities are being assessed by the algorithm. This study examines how people construct “snap” folk theories about the ways personality algorithms operate as well as how they react when shown their outputs. Through qualitative interviews (n = 37) with people after being presented with their own profile, this study identifies a series of folk theories that people came up with to explain the personality algorithm across four dimensions (data source, scope, collection process, and outputs). In addition, this study examined how those folk theories contributed to certain reactions, fears, and justifications people had about the algorithm. This study builds on our theoretical understanding of folk theory literature as well as certain limitations of algorithmic transparency/sovereignty when these types of inferential and predictive algorithms get coupled with people’s hopes and fears about employment, hiring, and promotion.

Keywords
algorithms, personality, folk theories, employment

Introduction
At its most basic, an algorithm is a syntax which embodies a command structure and provides “a sequence of instructions telling a computer what to do” (Domingos, 2015, p. 1). From a computer science perspective, whether it is matching search queries, curating social media feeds, or assembling meaningfulness in information flows, algorithms are seen as a technical solution to a technical problem—the formalization of defined strategies to categorize and sort information (Blass & Gurevich, 2004). Social scientists, on the contrary, have tried to broaden the realm of algorithmic inquiry, noting that they only stay as “inert, meaningless machines until paired with databases upon which to function” (Gillespie, 2014, p. 169). Once they begin certifying knowledge, producing algorithmic publics, and making predictive inferences, their importance moves beyond the technical to the social, as their design, values, and biases can have grave material consequences (Bucher, 2018; Eubanks, 2018; Noble, 2018). Studying them only as a material artifact overlooks how they are mediators of human knowledge and social experience (Gillespie, 2014), especially as they become more voluminous in their data collection, personalized, and predictive (Kitchin, 2017; Matz et al., 2017; Reviglio & Agosti, 2020; Zuboff, 2019).

The gap between the technological and social understandings of algorithms is sometimes identified as black box that needs to be unpacked and explored (Diakopoulos, 2015; Pasquale, 2015). Oftentimes, corporate strategy/design contributes to the lack of algorithmic transparency (Burrell, 2016). Transparency alone is insufficient, however, because
it is illegible to most non-technical users and solely descriptive (Ananny & Crawford, 2018; Zarsky, 2016). Visibility is another issue, as people often only see the algorithm at work through its outputs (Bucher, 2012; Petre et al., 2019; Sandvig, 2015). They may only become aware of issues with the algorithms in moments of controversy or as a result of backlash to algorithmic changes (Gillespie, 2014; Noble, 2018).

Algorithms are obscured by several forms of opacity, including lack of technical transparency, user illiteracy, and the characteristics and execution of machine learning algorithms themselves (Burrell, 2016).

One promising avenue of research for bridging the divide has been to examine users’ algorithmic “folk theories,” defined as a person’s intuitive, causal explanation about a system that guides their thoughts, beliefs, and actions with that system (Gelman & Legare, 2011). Recent studies have examined folk theories that people give for how they think an algorithm works (DeVito et al., 2017; Eslami et al., 2016; French & Hancock, 2017; Siles et al., 2020; Toff & Nielsen, 2018). Because a one-to-one computational understanding of algorithms is obscured, folk theories become an important mediator of people’s attitudes, perceptions, and behavioral responses to algorithms. Even though some folk theories are ultimately illogical and inaccurate, their association with certain affective responses and practices are important to understand (Bucher, 2018).

While algorithmic folk theory research is a growing area, much of the existing work has focused on social media feeds (DeVito et al., 2017; Eslami et al., 2016; Rader & Gray, 2015), video and streaming sites (Siles et al., 2019), search engines (Toff & Nielsen, 2018), and music recommendation services (Siles et al., 2020). Research on emerging algorithmically generated personality profiles, which pull from multiple sources and make direct statements and predictions about a person, have yet to be explored through a folk theory lens. Folk theory research has also primarily been conducted on known platforms, where users have at least some knowledge and experience of the algorithm. There has been less work on spontaneous or “snap” folk theories, where users are confronted with an unknown algorithm.

This study addresses these questions through an examination of Crystal Knows, a company that uses algorithms to produce an application that they claim can “tell you anyone’s personality.” We explore the relationship between folk theories about these personality recognition algorithms and affective responses by showing people their real profiles and conducting qualitative interviews. By identifying and linking a set of folk theories to particular responses, this study deepens our theoretical and empirical understanding of folk theories and how they are related to specific rationales, attitudes, and behavioral intentions. Given the increasing sophistication of personality algorithms, the prevalence of social media use, and the potential consequences of these algorithms, it is important to understand how people theorize and respond to these algorithms.

Figure 1. Number of papers on personality in IEEE Xplore and ACM Digital Library. Source. Vinciarelli and Mohammadi (2014).

**Automatic Personality Recognition Algorithms**

Personality testing is an estimated $500 million (USD) industry (Harrell, 2017). Organizations and managers utilize them to make hiring decisions, form teams, and assess performance (Moscoso & Salgado, 2004). However, there have been concerns about the efficacy of personality tests, as well as fears that people could manipulate responses to fake certain personality traits (Birkeland et al., 2006). There are also legal limits to who can take them—U.S. courts have ruled employer mandates for personality tests may discriminate against certain persons with disabilities.¹ They can also be limited in their access to certain populations outside of their organization.

Due to these constraints, there has been a turn toward computational psychometrics, a growing field that works on automatic personality recognition algorithms, defined as “the task of inferring self-assessed personalities from machine detectable distal cue” (Vinciarelli & Mohammadi, 2014, pp. 4–5). These computational algorithms attempt to detect personality based on posts, likes, and other metadata online (Bleidorn & Hopwood, 2019; Kosinski et al., 2013; Quercia et al., 2011). One systematic review found that the number of papers in computing databases that study personality has been steadily growing (Figure 1).

These tend to primarily focus on computational questions, such as how to conceptualize personality, which operational measures to use (e.g., Myers-Briggs, Eysenck Personality Questionnaire), and how to test ecological validity of personality assessments outside of self-report (Vinciarelli & Mohammadi, 2014). The ultimate goal is computational “accuracy,” and studies have touted computer-based personality judgments as being more accurate than human judgments (Youyou et al., 2015). Computer scientists also investigate whether all types of personality traits can be predicted simultaneously, which behaviors are most predictive, how gender and age relate to personality, and how much accuracy is lost by porting social media metrics across platforms (Bleidorn & Hopwood, 2019; Kosinski et al., 2013; Matz et al., 2017; Ortigosa et al., 2014; Schwartz et al., 2013).

¹Persons with disabilities.
Table 1. Folk Theories About Algorithms.

| Folk theory                                | Rationale                                                                 |
|--------------------------------------------|---------------------------------------------------------------------------|
| Personal Engagement Theory                 | The more interactions you have with someone, the more likely their content is to appear |
| Global Popularity Theory                   | The more likes something gets, the more likely it is to appear             |
| Format Theory                              | Certain types of stories would appear more (e.g., text over videos)       |
| Narcissus Theory                           | The more similar the poster was to them, the more likely it was to appear  |
| Theory of Loud and Quiet Friends           | Frequency of posts by the sender influences how likely their post is to appear |
| Eye of Providence Theory                   | The Social Network is all-seeing and curating each individual post based on specific preferences and topics |
| The Original Content Theory                | Original posts are more likely to be shown                                |
| Fresh Blood Theory                         | Posts from new/recent friends are more likely to be shown                 |
| Randomness Theory                          | No pattern to the algorithm                                               |

Source. Adapted from Eslami et al. (2016).

Some of the earliest work on reactions to personality algorithms found that they were seen as “creepily accurate,” although people also reported high levels of trust in the algorithm (Warshaw et al., 2015). Even though most people saw personality information as quite sensitive, they also felt pressure to publicize them (Gou et al., 2014; Warshaw et al., 2015). Context plays a role, with a majority stating they would be comfortable allowing employers to view their personality profiles (Gou et al., 2014). These paradoxical attitudes toward personality profiles are an important area for research, as there has also been recent work that explores how people try to circumvent and trick personality recognition tools in a simulated conversation (Völkel et al., 2020). In addition, early work assessing participant responses to personality algorithms was conducted in laboratory settings (Graves & Matz, 2018). Crystal Knows is a live database that is generating profiles about real people. This opens up new avenues for research beyond hypothetical attitudes about people’s specific perceptions, attitudes, and folk theories about how Crystal operates (Bivens, 2015).

Folk Theories and Responses to Algorithms

Folk theories have long been an important part of understanding human–computer interaction, as their reasoning about technology informs their expectations, interactions, and practices surrounding a technology (Norman, 1988). Their application to algorithms has been instructive, given (1) that the intuitive nature of folk theories can surface understandings that are unconscious, implicit beliefs about a system; (2) they can describe topical, explanatory, and causal relationships; and (3) they are the product of descriptions about and lived experience with algorithms (French & Hancock, 2017). They are simultaneously an object of study, an implicit collection of beliefs about the system, and an analytical frame to understand and rationalize people’s behaviors surrounding algorithms (DeVito et al., 2017; Rader & Gray, 2015).

Eslami and colleagues (2016) were some of the first to map folk theories about personalized social media feeds, as people offered several theories explaining why certain stories would show up and what they did in response (Table 1).

Some folk theories have also surfaced when certain companies made changes to their algorithms, such as moving from a time-based to a curated feed (DeVito et al., 2017; Rader & Gray, 2015; Skrubbeltrang et al., 2017). They can also vary depending on what sources of information people draw upon when constructing these folk theories (DeVito et al., 2018; French & Hancock, 2017).

Beyond identifying folk theories that people hold, they have also been used to explain people’s awareness, attitudes, and behaviors surrounding algorithms. As algorithms become entangled with practice, people will try to test them and learn about them to formulate folk theories, and perhaps even adjust their behaviors around them (Eslami et al., 2016; Gillespie, 2014; Sandvig, 2015). Folk theories have been found to affect people’s perceptions of algorithmic fairness and trust (Lee, 2018). Some studies theorize that confronting people with stories about data flows can change people’s willingness to disclose and modify their privacy settings (Nosko et al., 2012). Others have found relationships between algorithmic folk theories and certain practices of self-presentation (DeVito et al., 2018), perceived relational closeness (Eslami et al., 2015), behavioral responses such as hiding posts (Eslami et al., 2016), and other “users performing protective tactics and deliberate attempts to influence or circumvent algorithmic systems through interactions with them” (Lomborg & Kapsch, 2020, p. 755).

In contrast to people taking protective actions based on higher levels of awareness, one commonly identified phenomenon that scholars have found with tracking/algorithms is the privacy paradox, which refers to a disconnect between stated attitudes about privacy and actual protection behaviors (Acquisti & Gross, 2006; Debatin et al., 2009). One explanation may be that the social, economic, and culture pressures toward utilizing these algorithms simply outweigh the costs of quitting (Baumer et al., 2013; Warshaw et al., 2015). Other scholars have found that instead of motivating change, stories of continual algorithmic intrusions can lead to “digital resignation,” where people
get worn down, adjust their expectations, and accept algorithms as the new normal (Draper & Turow, 2019). Another potential response is “psychological hedonism,” whereby users see the algorithm as gratifying, benevolent, and potentially beneficial to their standing (Gal, 2018). Even within the same system, studies have found that people can hold two very different folk theories (e.g., personification of an algorithm compared to computational machine), which leads to a multidimensional set of responses (Siles et al., 2020).

This study builds on folk theory literature through the case of automated personality recognition algorithms in several ways. First, personality algorithms are an emerging context, with different personal outputs and evidence of their inner workings than social media feeds or music recommendation sites. While there may be similarities to existing folk theories (Eslami et al., 2016; French & Hancock, 2017; Woodruff et al., 2018), the nature of personality algorithms vary in origin, scope, outputs, and specificity. Second, most existing studies assume a level of algorithmic awareness and have been conducted on platforms where users have some agency in their participation and a level of familiarity with the platform (Hamilton et al., 2014). That familiarity can shape assessments of how the algorithm is functioning and why it is showing certain links/stories/tweets/songs (Eslami et al., 2016; Siles et al., 2020). There has been less work done about unexpected “snap” folk theories that emerge when the algorithm and platform are totally unknown, the users do not know why they were selected, and are just finding out they have been algorithmically sorted. When confronted with an unknown algorithm, folk theories may play an even larger role in people’s assessments than prior experience, descriptions of the algorithm, or trust in the company. Third, personality algorithms such as Crystal may be seen as particularly intrusive, because it goes beyond moderating a feed to making second-order predictive statements specific to an individual (Bivens, 2015; Gou et al., 2014; Warshaw et al., 2015). Applying folk theory research to these algorithms can help assess people’s responses while also broadening our understanding of the relationship between folk theories and specific reactions. Hence, we ask the following research questions:

RQ1. When presented with their own algorithmically generated personality profile, what folk theories do people construct for how the algorithm works?

RQ2. Based on the folk theories they hold, what types of rationales and responses do people report about the algorithm?

Methods
Case Application
Crystal Knows is a company based in Nashville, founded in 2015. According to the company, they “aggregate public data, analyze the text content, detect communication style and personality assumptions, and show this information to everyone who uses Crystal.” This profile includes the person’s name, photo, a set of personality assessments, and recommendations for how to communicate and interact with this person. Monthly subscriptions to their database of personality profiles are sold to advertisers, prospective employers, and everyday people for networking, hiring, personnel, and team building. Besides specific personality assessments of people through graphs and the DISC model (Dominance, Influence, Steadiness, Conscientiousness), they also offer recommendations for how to communicate with these individuals, how to convince/persuade them, whether they would be an appropriate candidate for certain positions, and how well they would work with others.

Recruitment and Procedure
Participants were recruited in undergraduate classes between Fall 2018 and Spring 2019 at a university in the Midwest United States. Eligibility criteria included being 18 years or older, English speaking, and already having a Crystal Knows profile. Once students volunteered, their name was run through the Crystal Knows database to see if there was a profile. For people with profiles, interviews were scheduled and conducted. The authors have no financial relationship with Crystal as a company, and participants were told this as part of obtaining written consent. Upon completion, they had a choice of receiving extra credit in the course they were recruited from or a $10 gift card.

Participants were first shown a printed copy of their Crystal Knows profile and told they could examine it for as long as they wanted to, and to make written notes in places where they thought certain recommendations were particularly notable or interesting. They then proceeded to the interview portion of the study. Because every person’s profile varied in terms of their recommendations and what they found notable, semi-structured interviews were conducted with an initial set of questions and follow-ups based on participant responses. Qualitative interviews are also found to be particularly useful when paired with algorithmic probes for understanding people’s rationales, reactions, and folk theories (Eslami et al., 2016; Woodruff et al., 2018). Interview questions started with their initial notes and reactions, and then moved on to ask about how they felt they were portrayed in the profile, their feelings about the algorithmic production, folk theories they had about how it was generated (often without their knowledge/consent), circumstances where they may or may not want their profile to be used, and whether they would alter their online practices as a result of seeing their algorithmically generated profile.

There was a total of 37 participants. Of those, 27 participants identified as women. Of our sample, 73% identified as Caucasian, 14% identified as Black or African American, 11% identified as Asian, and 3% identified as multi-racial. Interviews lasted between 30 and 87 min ($M = 50$).
Data Analysis

Interviews were audio-recorded, transcribed, and analyzed in Dedoose. There were two stages of analysis, the first was for RQ1 to identify folk theories that participants had about Crystal. We identified four primary domains where people constructed folk theories about the Crystal algorithm—(1) Data Source, (2) Data Scope, (3) Collection Procedures, and (4) Personalization of Output. Along each dimension, we categorized folk theories that posited a more limited/basic/simple algorithm, as well as those that theorized a powerful/complex/personalized algorithm.

For RQ2, the second part of the data analysis was to code for people’s responses to the algorithm, based on the folk theories they held. We found several themes related to fear of the algorithm, fear of algorithmic consequences, pride/admiration/liking, and blame/resignation. We then linked these responses to the folk theories participants reported about the algorithm itself, to see if there were common patterns of reactions based on certain folk theories. By linking folk theories to participant rationales and explanations of behavioral intentions, we are able to map certain pathways for responses that form the basis of our findings.

Findings

Across multiple domains of how the algorithm functions, respondents came up with a wide range of folk theories. They often fell along a spectrum based on what they were seeing, with some believing that the algorithm was basic, limited, and constrained and others believing the algorithm was all powerful, complex, and multidimensional.

Data Source—Single Origin (Limited) to Aggregated Origins (Powerful)

Respondents gravitated toward profile cues that pointed to how they were identified and the source of the data. Some of the clearest indicators was the picture associated with their profile, the profile name, particularly if they use different nicknames or abbreviations across social media platforms: “I do find it fascinating because my real name isn’t Ellie, it’s Elizabeth.” Others thought that it was based on information they volunteered through other personality tests: “I’ve taken these personality tests before, so they pull stuff off of those” (Anna, F). These they took to be indications of a single data source.

Some people either overlooked the picture or presumed that it was just one source of many: “How would they know through my social media that that’s something I need to do? […] I would assume like maybe LinkedIn and Instagram. […] It’s like they’re looking at my email” (Grace, F). Many respondents speculated about whether Crystal had access to information beyond social media platforms, such as email, text messaging, phone records, deleted posts, photo scanning, and/or location:

Their secret file is probably like my locations of where I posted some stuff and where I typically would be on those social media platforms because . . . they ask you “are you willing to share your location.” (Linn, F)

People knew it was coming from somewhere, but they differed in whether they thought the data were coming from one source or across multiple platforms.

Data Scope—Snapshot Theory (Limited) to Timeline Theory (Powerful)

There were two competing theories as to the scope of data the algorithm was working with. The snapshot theory assumes that the algorithm works based on current information when the company decides to enter you into the database:

I think it would show a time […] that this is how they were at this point. […] That’s like going more into detail but again, this is more like an overview in a way and it’s not like it’s pulling up my baby pictures that my mom posted of me or like my 13th birthday or stuff like that. (Linn, F)

Conversely, the timeline theory assumes that the algorithm has been tracking them from whenever they started using social media: “It must have been awhile. I would assume ever since I—when did this come out—2014, it feels like they’ve probably been doing this since 2014 then” (Anna, F). The timeline theory also believes the algorithm is continually learning and inputting information from your posts to change their assessments:

Personality changes. […] If they were to do this for me in 10 years, […] there would be some core things that would definitely be the same, but I think they would have to be sensitive to that. I don’t think it’s something […] the algorithm would have to look at every day because obviously, that’s not going to change, but I think […] just gathering more information, the more you post, this is going to change this stuff too. (Emma, F)

Collection Procedures—Settings Protection Theory (Limited) to Weighted Personalization Theory (Powerful)

Participants also considered the precise nature of the algorithm itself, what exactly it was processing, and how it was doing so. Settings protection theory posits that Crystal is abiding by certain data privacy settings, and that the algorithm could only be pulling data from social media platforms where they had set it to public:

So Facebook I have my privacy settings you can see my profile pictures and maybe a few tag pictures, but you can’t see my posts if you’re not friends with me. […] Twitter is I’m public on that. You can see anything. […] So maybe they found from tweets my last name in it. (Hannah, F)
By preventing the algorithm from accessing certain data, that could limit its ability to perform certain calculations: “I feel like if you have a Facebook profile, how strict your privacy settings have to be means it actually doesn’t read stuff” (Sasha, F).

Weighted Personalization theory, on the contrary, posits that Crystal has access to all or several sources of information, and that it weights the personality profile toward the platform where there is the most activity and tangible outputs:

Well, that’s where I think I post most. And, well, I post most on Facebook and the most stuff that I post on Facebook is my personal opinion . . . So I think Facebook is where I use the most words and maybe they pulled from that. (Evelyn, F)

It also assumes the algorithm is powerful enough to make determinations as to which platforms and posts are more revealing of personal opinion or personality: “I’m assuming from context clues and my own knowledge that the first thing listed is probably the strongest one, compared to this one, this is an on-going list that’s pulled” (Linn, F).

Personalization of Output—Astrology (Limited) to Omniscience (Powerful)

Accuracy of their profiles was one method by which people tried to gauge personalization, with some people feeling that the algorithm was almost perfect (95%–100%) while others were more skeptical (70%–80%). Certain statements that they felt were inaccurate, combined with the general nature of some of the profile statements, led some to question the specificity and power of the algorithm: “I think it’s very basic. […] I know there was some different things . . . But I don’t see it as like a knowing all type of personality. It’s like astrology.” (Ariel, F).

Others pointed to specific statements that led them to conclude it was highly personalized:

I was a little skeptical, just because I feel like […] a lot of horoscopes everyone can be like oh, I can relate to that. […] But the fact that mostly everything is like, on point, is a little more convincing […] Won’t give you a straight “no,” […] Expect to lead the conversation. I feel like a lot of people are extroverted, and they lead the conversation. So not everyone can relate to that. (Leah, F)

Often it was one specific statement that would cause participants to move more toward omniscience theory, that the algorithm was all-seeing and all powerful across platforms:

I would get thrown off by one comment. Like don’t be too formal is on there […] I don’t know if things like this can read your texts or emails, but like I email a ton for business and stuff. So, maybe that. […] it’s very complex I doubt you could surprise me at all if you told me they could access that. (Sasha, F)

Perceived accuracy often led people to conclude that the algorithm was generating individually specific content:

These sections “when speaking to you,” “what drains you,” […] I don’t know how they’d get that information […] on me, and they’re able to extract so much from that information, I feel like you can’t even speak for yourself when something like this is out here, so makes me a little worried. (Nathan, M)

Folk Theories and Affective Responses

While some people uniformly subscribed to either limited or complex folk theories about the algorithms, it is also possible to hold limited theories for some aspects of the algorithm and complex ones for the others. For example, people might believe in a powerful algorithm that drew on various data sources and tracked them over a long period of time, but conclude that the recommendations are inaccurate or general. Conversely, it is possible to think that the algorithm is based on a single origin snapshot of publicly shared data, but feel as if the recommendations are quite personalized. Because of this, we were able to analyze how certain folk theories and combinations of folk theories corresponded with certain responses, as well as their rationales for reaching that conclusion.

Fear of the Algorithm

People who held limited folk theories about the algorithm’s data source, scope, collection procedures, and specificity tended to report less fear of the algorithm itself, for a variety of reasons. The limited folk theories about single origin sources, snapshot view, and astrology theories allowed people to dismiss it as less threatening: “For me, it’s so generic, or general, that it’s just if I had the suspicion this was coming from somewhere more specifically that was really close, then yeah, I’d be uncomfortable. But I can’t pinpoint this to anything” (Theo, M). They also saw it as more lighthearted and fun, to be viewed like an online personality quiz: “It could kind of be fun. […] I like taking tests about personalities because I don’t know how I am. It’s okay because that’s public, when you’re posting stuff that people are gonna have access to it” (Addy, F).

People who viewed the algorithm as more powerful tended to see it as more intrusive and extractive:

Because I feel like […] it just like takes, […] It is just creepy that it’s an algorithm that did all this, and not like a person, it’s just a computer. […] And it’s just not good to be sorted. (Jeanine, F)

People would be particularly upset if the algorithm gave indications that it was not acting according to settings protection theory, or otherwise gaining consent: “That’s my LinkedIn picture. That’s kind of creepy, honestly. Because I’m private on every social media so it’s weird that they might not be private” (Chiara, F).
Fear of Algorithmic Consequences

All respondents, regardless of whether they held limited or powerful folk theories, talked about fear of consequences. Most pressing was the employment context, because the Crystal profiles are geared toward hiring purposes and the demographic of the users: “I just don’t like how it’s so much, it’s all in the computer. [. . .] It takes a lot of thinking out from the company’s standpoint; the person who’s going over it, the hiring manager, and all of that” (Anna, F). The nature of the fears, however, were slightly different depending on the type of folk theory.

People who held limited folk theories about the algorithm were primarily worried about issues arising from inaccuracy: “Like 75–78% accurate. [. . .] I think the main thing that they got wrong, I am pretty quiet. [. . .] I have to make it 100% accurate, yeah” (Grace, F). Even if they believed that the algorithm was not particularly sophisticated, the fear of negative consequences was enough to push them toward trying to correct their profile: “I would love to take the test and get it even more accurate. [. . .] I just want it to be 100% accurate so I can send this off to my employers” (Nathan, M).

People who held powerful theories were similarly concerned, but more that the extensive tracking of the algorithm could reveal certain things that might jeopardize their chances:

They could simply pull this up and be like, “I don’t like that she has this, this and this.” Like yeah I don’t like confrontation, and I don’t like change. If an employer knew that and might not get . . . if that wouldn’t get me hired that’s all upsetting. Because I’m more than just these four pages. (Nina, F)

People in disadvantaged groups also saw cause for concern with regard to discrimination, that the algorithm could be making predictive assessments that are coded with gender and racial stereotypes: “Aggressiveness, detachment, those kinds of things. Race wise, the main stereotypes at least for black women would be more angry” (Shannon, F). Others saw their profiles confirming gender roles:

Like people can easily walk over me, I am gentle. [. . .] She doesn’t like loud language. Overly polite. Not a leader in any way [. . .] very like emotionally sensitive, which is like stereotypical woman checking on friends. (Nina, F)

Pride, Affinity, and Admiration

One response that was specific to people with powerful folk theories was a feeling of pride, that they were important enough for the algorithm to catalog and that someone was picking up on things they were putting out into the world: “I’m an avid [social media] user and always have been. [. . .] I knew I’d be on there because of me being so easily found” (Gillian, F). Not only were they happy to be noticed by the algorithm, they also expressed a grudging admiration for it because of its perceived accuracy:

I would be more bothered if they weren’t so accurate. [. . .] I know it’s a thing that’s probably going to happen, so I can’t be mad about it. [. . .] So for me [. . .] I would see this as a benefit. (Shannon, F)

While they agreed that the lack of transparency was unsettling, the outputs justified the means: “I think it’s just more creepy how someone made this profile about me and then put this stuff in that. It’s creepy but cool.” Some even expressed strong preferences for this type of system as advantageous:

Because Crystal’s really just reflecting back on you, and if you don’t get hired for a job because of Crystal, then maybe I just wasn’t actually meant to have the job. [. . .] I can just look at this and I’ll be like, “This is what the system says.” [. . .] The computer knows what they want. I’m just going to trust the computer. This is objective because it’s an outside eye. (Yin, F)

Blame, Protection, and Resignation

Regardless of theories about data source, the participants were captured by the Crystal algorithm. As a result, most people were adamant that the blame fell on the people for posting, rather than the algorithm processing that information:

I think if you have enough of a social media presence to where it can do it [. . .] you’re asking for it. But if you have that kind of online presence then I don’t think it’s an invasion of privacy, because you’re putting yourself out there. (Kayla, F)

Often, they blamed themselves too, that they likely signed up for something inadvertently: “It is kind of weird that there’s this profile on the website I didn’t know existed. [. . .] But at the same time I’m guessing is pulling from things that I gave permission to posting on the Internet” (Lacy, F).

Ideas about protecting themselves from the Crystal algorithm varied depending on people’s theories about the scope of data and collection processes. Some wanted specific social media feeds to be removed from the algorithm itself and to delete posts, if they believed that the algorithm was tracking things historically: “I might ask to have that Twitter deleted [. . .] Again, I thought I had most of those private” (Kylie, F). Others who believed in settings protection theory noted that Crystal was motivating them to revisit and change their settings on certain profiles: “I’m starting to look at my privacy settings a little more thoroughly. This just another thing adding on to why I should look at the privacy settings” (Evelyn, F).

People holding powerful folk theories also reported a sense of resignation, that nothing they did mattered including changing their settings:

The answer is gonna be no. I’m still gonna continue to like the same things I like and follow the same accounts and accept the accounts. But if they tell you [. . .] about Crystal. What? Personality? No! Then it’s like, “Okay, well you can’t have an account then.” They’re like, “We’re gonna take everything.”
Crystal was just more evidence that these encroachments were inevitable: “I had no idea about it, and my whole class is like ‘What in the world?’ [. . .] To people it’s like ‘Oh, Facebook data breach. Shrug. Like . . . Equifax. Shrug’” (Dylan, M). The range of folk theories across these various algorithmic domains, as well as the relationship between specific folk theories and responses, is visualized in Figure 2.

**Discussion**

This study is one of the first to examine participant responses to an active, publicly available personality recognition algorithm outside a laboratory setting (Graves & Matz, 2018). Mapping the various dimensions of folk theories as well as specific rationales people have to personality algorithms is an important first step toward explaining and clarifying why people might hold counterintuitive or paradoxical responses (Figure 2). The framework documents the folk theories people formulated across various levels of algorithmic function, how certain theories veered more toward a limited view or a powerful view, and the specific ways that they informed attitudinal and emotional responses.

One of the first questions that people had was the data source. While some participants looked to tangible cues (e.g., name, photo) to identify a single origin, others imagined that the algorithm had access to a wide range of information from multiple accounts, text/emails, location, and so on. While existing studies have found that folk theories are generated by identifying multiple sources of data (DeVito et al., 2018), with Crystal, they had to discern what personal content among a myriad of online activities the algorithm was drawing from. In the end, the precise nature of the folk theory was less important than the dimension of the folk theory in predicting a response. In both limited and powerful readings, regardless of how in-depth they believed the sourcing to be, people blamed themselves for being captured by the algorithm.

In addition, our findings identified folk theories across multiple dimensions of the algorithm. Much of the existing literature on folk theories have been focused on one dimension (e.g., algorithmic process) and how people believe a particular feed functions (DeVito et al., 2017; Rader & Gray, 2015; Skrubbeltrang et al., 2017). This study, by examining an unknown algorithm, allows us to identify folk theories on multiple levels of algorithmic function and how certain theories can override others. In seeing how people were more willing to assign self-blame for their posting behavior, we can better understand one mechanism by which people rationalize algorithmic objectivity and view them as a “Transparent Platform” that is unfiltered and serves as a mirror reflecting human action (French & Hancock, 2017; Gillespie, 2014). This suggests that folk theories at the initial level (e.g., data source) can override other folk theories in terms of responsibility, where the choice to participate in social media at all supersedes any harms that befall you.

At other levels, however, people still did try to figure out how the algorithm operated. While much of the existing algorithm and tracking literature has been trying to understand and motivate people to take protective actions (Acquisti & Gross, 2006; Debatin et al., 2009; Nosko et al., 2012), this study found that these responses are mediated by folk theories. Different folk theories about source, scope, and collection procedures are related to the perceived effectiveness of certain actions as well as people’s sense of efficacy.

Clarifying the relationship between specific folk theories and affective responses to personality algorithms is a key contribution of this study. While existing research into social media feeds has identified various different folk theories (Eslami et al., 2016; French & Hancock, 2017), the relationship between folk theories and responses varies dramatically depending on the platform (DeVito et al., 2018). Some have also found that folk theories sometimes compete or are mutually exclusive (Eslami et al., 2016; Siles et al., 2020). By mapping out two threads of folk theories across domains, this study begins to understand how a set of folk theories is linked to particular responses. Folk theories can also explain how people may have the same reported emotional and affective response, but very different rationales as to why they feel that way.

For example, while people who held both limited and powerful folk theories reported fear about the algorithm, the nature of their fears were different. People who thought the
algorithm was quite limited mostly feared inaccuracy, while people who believed the algorithm was powerful feared too much accuracy. The latter was particularly concerning, because Crystal targets industries that could potentially use these algorithms to evaluate, categorize, and potentially discriminate against them (Bivens, 2015; Matz et al., 2017). Just as predictive algorithms have proved problematic in policing and government services (Eubanks, 2018), personality algorithms could amplify that risk beyond exposure and into power-laden situations where people are making decisions. Particularly as companies attempt to identify negative personality traits that could predict poor performance (Moscoso & Salgado, 2004) or resignations (Silverman & Waller, 2015), the push toward algorithmic personality-based hiring, promotion, and evaluation raises important questions of fairness which could affect people’s trust in the companies that utilize them (Lee, 2018; Woodruff et al., 2018).

Although fear was a prevalent response, another way that people who held powerful theories about the algorithm responded was to say positive things about the algorithm (e.g., pride, affinity, admiration). While people holding limited theories could dismiss the algorithm as unimportant or inaccurate, people holding powerful theories may be susceptible to psychological hedonism for several reasons: (1) being in the Crystal database was viewed with some pride (e.g., they are important enough to catalog), (2) the profiles present individualized personality statements, and (3) they are mostly positive and reflect a persona which people want to be true. Hence, people report being pleased with the algorithm because their specific profile was largely seen as benevolent, which could make them less sympathetic to others (Gal, 2018). With a belief that the algorithm is omniscient and all powerful, people feel as if a positive profile is evidence of their exemplary behavior. If other people’s profiles are negative or inaccurate, then it must be their fault either for the content of their social media activities or even just making the decision to be active on social media.

Importantly, while existing studies have tried to raise awareness about algorithms (Hamilton et al., 2014) and increase their fear of them (Nosko et al., 2012), this study finds that certain folk theories may explain seemingly counterintuitive responses to increased awareness. For example, while many people who held powerful theories about the algorithm reported fear, thinking the algorithms were powerful also inspired admiration and affinity for the algorithm, as well as digital resignation (Draper & Turow, 2019; Zuboff, 2019). Personality algorithms geared toward the work force may also leverage economic incentives, because employment seeking people want to be seen and want to be seen in a positive light. While algorithmic sovereignty has been proposed to give people control over their data and who can use it (Revgilio & Agosti, 2020), sovereignty may be weakened by systems like Crystal that use external economic pressure to get people to volunteer their data. In addition, the desire to be seen and the “threat of invisibility” could have material consequences on employment, which could further discipline people into participating (Bucher, 2018). Even in the face of an algorithm that pulls an unknown amount of data from opaque sources, does so without consent, and produces a profile that is intended for other contexts where the user is at a power disadvantage, many concluded that the solution is simply to “get friendly with bots,” with the way forward being to “build, maintain, and improve upon code and algorithms” (Steiner, 2012, p. 218).

Finally, because Crystal was an unknown algorithm that people were confronted with, this study builds on our understanding of “snap” folk theories, which are distinct from familiar theories built over time (DeVito et al., 2017; Eslami et al., 2016). First, snap theories are more susceptible to algorithmic cues, as opposed to experientially derived. This makes the presentation of cues from the algorithm itself especially salient (Xu & Liao, 2020). Second, snap theories are formed about the algorithm, company, and outputs, all in a short period of time, which dictates people’s future interactions with the technology. As people continue to engage with these technologies, these snap theories can fade over time or “harden” into more concrete theories through confirmatory biases. Understanding the relationship between snap folk theories and responses can be especially useful in guiding efforts toward change, particularly in terms of how they solidify into beliefs and assessments of fairness (Lee, 2018; Zarsky, 2016). Existing research has found that there is significant room for movement: “the majority (60.7%) of participant folk theories were malleable, suggesting that most folk theories are flexible, as opposed to closely-held, rigid beliefs” (DeVito et al., 2018, p. 8). Understanding snap theories is an important step for future research, in terms of how people can gravitate toward initial theories as well as how more familiarity could potentially move them off of their snap theories, all of which can inform future efforts to mediate attitudes and behaviors.

**Conclusion**

Alongside ever evolving complexity in data collection systems, machine learning, and big data, automatic personality recognition algorithms are an emerging phenomenon that lies at the intersection of these developments. By revealing some initial folk theories and their relationships to certain responses to this system, this study lays the groundwork for future work in these areas. While this study explored people’s qualitative folk theories, rationales, and reactions, future quantitative research might be better suited to validate the prevalence and strength of certain pathways relationships. In addition, this study does not attempt to identify differences among groups for holding certain theories. More quantitative work into various demographic groups that respond to personality algorithms could further build on our existing framework and potentially identify stronger causal relationships. As Crystal and other automated personality recognition algorithms are making bold promises about their utility in fields such as employment, advertising, politics, persuasion, admissions, and online dating, they are an important emerging context for social media and algorithm researchers to explore.
Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Tony Liao https://orcid.org/0000-0003-1889-3831

Notes

1. Karraker v. Rent-A-Center, Inc., 411 F. 3d 831 (7th Cir. 2005).
2. Crystal Knows Privacy Policy, Frequently Asked Questions (2020).

References

Acquisti, A., & Gross, R. (2006, June). Imagined communities: Awareness, information sharing, and privacy on the Facebook. In International workshop on privacy enhancing technologies (pp. 36–58). Springer.

Ananny, M., & Crawford, K. (2018). Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. New Media & Society, 20(3), 973–989.

Baumer, E. P., Adams, P., Khovanskaya, V. D., Liao, T. C., Smith, M. E., Schwanda Sosik, V., & Williams, K. (2013, April). Limiting, leaving, and (re)lapsing: An exploration of Facebook non-use practices and experiences. In Proceedings of the SIGCHI conference on human factors in computing systems (pp. 3257–3266). Special Interest Group on Computer–Human Interaction.

Birkeland, S. A., Manson, T. M., Kisamore, J. L., Brannick, M. T., & Smith, M. A. (2006). A meta-analytic investigation of job applicant faking on personality measures. International Journal of Selection and Assessment, 14(4), 317–335.

Bivens, R. (2015, June 9). Interrogating Crystal’s design flaws highlights options for technical advocacy. Culture Digitally. https://culturedigitally.org/2015/06/interrogating-crystals-design-flaws-highlights-options-for-technical-advocacy/

Blass, A., & Gurevich, Y. (2004). Algorithms: A quest for absolute accuracy. In Current trends in theoretical computer science: The challenge of the new century—Vol. 1: Algorithms and complexity, Vol. 2: Formal models and semantics (pp. 283–311). https://www.microsoft.com/en-us/research/wp-content/uploads/2017/01/164.pdf

Bledorn, W., & Hopwood, C. J. (2019). Using machine learning to advance personality assessment and theory. Personality and Social Psychology Review, 23(2), 190–203.

Bucher, T. (2012). Want to be on the top? Algorithmic power and the threat of invisibility on Facebook. New Media & Society, 14(7), 1164–1180.

Bucher, T. (2018). If . . . then: Algorithmic power and politics. Oxford University Press.

Burrell, J. (2016). How the machine “thinks”: Understanding opacity in machine learning algorithms. Big Data & Society, 3(1), Article 622512.

Crystal (2020) Knows privacy policy FAQ. Available at: https://www.crystalknows.com/privacy

Debatin, B., Lovejoy, J. P., Horn, A. K., & Hughes, B. N. (2009). Facebook and online privacy: Attitudes, behaviors, and unintended consequences. Journal of Computer-Mediated Communication, 15(1), 83–108.

DeVito, M. A., Binholtz, J., Hancock, J. T., French, M., & Liu, S. (2018, April). How people form folk theories of social media feeds and what it means for how we study self-presentation. In Proceedings of the 2018 CHI conference on human factors in computing systems (pp. 1–12). Association for Computing Machinery.

DeVito, M. A., Gergle, D., & Binholtz, J. (2017, May). “Algorithms ruin everything”: #RIPTwitter, folk theories, and resistance to algorithmic change in social media. In Proceedings of the 2017 CHI conference on human factors in computing systems (pp. 3163–3174). Association for Computing Machinery.

Diakopoulos, N. (2015). Algorithmic accountability: Journalistic investigation of computational power structures. Digital Journalism, 3(3), 398–415.

Domingos, P. (2015). The master algorithm: How the quest for the ultimate learning machine will remake our world. Basic Books.

Draper, N. A., & Turow, J. (2019). The corporate cultivation of digital resignation. New Media & Society, 21(8), 1824–1839.

Eslami, M., Karahalios, K., Sandvig, C., Vaccaro, K., Rickman, A., Hamilton, K., & Kirlik, A. (2016, May). First I “like” it, then I hide it: Folk theories of social feeds. In Proceedings of the 2016 CHI conference on human factors in computing systems (pp. 2371–2382). Association for Computing Machinery.

Eslami, M., Rickman, A., Vaccaro, K., Aleyasen, A., Vuong, A., Karahalios, K., . . . Sandvig, C. (2015, April). “I always assumed that I wasn’t really that close to [her]”: Reasoning about invisible algorithms in news feeds. In Proceedings of the 33rd annual ACM conference on human factors in computing systems (pp. 153–162). Association for Computing Machinery.

Eubanks, V. (2018). Automating inequality: How high-tech tools profile, police, and punish the poor. St. Martin’s Press.

French, M., & Hancock, J. (2017). What’s the folk theory? Reasoning about cyber-social systems. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2910571

Gal, M. S. (2018). Algorithmic challenges to autonomous choice. Michigan Technology Law Review, 25, Article 59.

Gelman, S. A., & Legare, C. H. (2011). Concepts and folk theories. Annual Review of Anthropology, 40, 379–398.

Gillespie, T. (2014). The relevance of algorithms. In Media technologies: Essays on communication, materiality, and society. https://mitpress.upennpressscholarship.org/view/10.7551/mitpress/9780262525374.001.0001/upso-9780262525374-chapter-9

Gou, L., Zhou, M. X., & Yang, H. (2014, April). 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 (pp. 955–964). Special Interest Group on Computer–Human Interaction.

Graves, C., & Matz, S. (2018, May 2). What marketers should know about personality-based marketing. Harvard Business Review. https://hbr.org/2018/05/what-marketers-should-know-about-personality-based-marketing

Hamilton, K., Karahalios, K., Sandvig, C., & Eslami, M. (2014). A path to understanding the effects of algorithm awareness. In CHI’14 extended abstracts on human factors in computing systems (pp. 631–642). Association for Computing Machinery.

Harrell, E. (2017, March–April). A brief history of personality tests. Harvard Business Review. https://hbr.org/2017/03/a-brief-history-of-personality-tests
Kitchin, R. (2017). Thinking critically about and researching algorithms. *Information, Communication & Society*, 20(1), 14–29.

Kosinski, M., Stillwell, D., & Graepel, T. (2013). Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the National Academy of Sciences of the United States of America*, 110(15), 5802–5805.

Lee, M. K. (2018). Understanding perception of algorithmic decisions: Fairness, trust, and emotion in response to algorithmic management. *Big Data & Society*, 5(1), Article 756684.

Lomborg, S., & Kapsch, P. H. (2020). Decoding algorithms. *Media, Culture & Society*, 42(5), 745–761.

Matz, S. C., Kosinski, M., Nave, G., & Stillwell, D. J. (2017). Psychological targeting as an effective approach to digital mass persuasion. *Proceedings of the National Academy of Sciences of the United States of America*, 114(48), 12714–12719.

Moscoso, S., & Salgado, J. F. (2004). “Dark side” personality styles as predictors of task, contextual, and job performance. *International Journal of Selection and Assessment*, 12(4), 356–362.

Noble, S. U. (2018). *Algorithms of oppression: How search engines reinforce racism*. New York University Press.

Norman, D. A. (1988). *The psychology of everyday things*. Basic Books.

Nosko, A., Wood, E., Kenney, M., Archer, K., De Pasquale, D., Molema, S., & Zivcakova, L. (2012). Examining priming and gender as a means to reduce risk in a social networking context: Can stories change disclosure and privacy setting use when personal profiles are constructed? *Computers in Human Behavior*, 28(6), 2067–2074.

Nobre, S. U. (2018). Algorithms of oppression: How search engines reinforce racism. New York University Press.

Reviglio, U., & Agosti, C. (2020). Thinking outside the Black-box: The case for “algorithmic sovereignty” in social media. *Social Media + Society*, 6(2), Article 915613.

Sandvig, C. (2015). Seeing the sort: The aesthetic and industrial defense of “the algorithm.” *Media-N*, 11(1): 35–51.

Schwartz, H. A., Eichstaedt, J. C., Kern, M. L., Dziurzynski, L., Ramones, S. M., Agrawal, M., . . . Ungar, L. H. (2013). Personality, gender, and age in the language of social media: The open-vocabulary approach. *PLOS ONE*, 8(9), Article e73791.

Siles, I., Espinoza-Rojas, J., Naranjo, A., & Tristán, M. F. (2019). The mutual domestication of users and algorithmic recommendations on Netflix. *Communication, Culture & Critique*, 12(4), 499–518.

Siles, I., Segura-Castillo, A., Solís, R., & Sancho, M. (2020). Folk theories of algorithmic recommendations on Spotify: Enacting data assemblages in the global South. *Big Data & Society*, 7(1), Article 923377.

Silverman, R. E., & Waller, N. (2015, March 13). The algorithm that tells the boss who might quit. *Wall Street Journal*. https://www.wsj.com/articles/the-algorithm-that-tells-the-boss-who-might-quit-1426287935

Skrubbeltrang, M. M., Grunnet, J., & Tarp, N. T. (2017). # RIPINSTAGRAM: Examining user’s counter-narratives opposing the introduction of algorithmic personalization on Instagram. *First Monday*. https://firstmonday.org/ojs/index.php/fm/article/view/7574/6095

Steiner, C. (2012). Automate this: How algorithms took over our markets, our jobs, and the world. Portfolio.

Tuff, B., & Nielsen, R. K. (2018). “I just Google it”: Folk theories of distributed discovery. *Journal of Communication*, 68(3), 636–657.

Vinciarelli, A., & Mohammadi, G. (2014). A survey of personality computing. *IEEE Transactions on Affective Computing*, 5(3), 273–291.

Vökel, S. T., Haeuslenschmid, R., Werner, A., Hussmann, H., & Butz, A. (2020, April). How to Trick AI: Users’ strategies for protecting themselves from automatic personality assessment. In *Proceedings of the 2020 CHI conference on human factors in computing systems* (pp. 1–15). Association for Computing Machinery.

Warshaw, J., Matthews, T., Whittaker, S., Kau, C., Bengualid, M., & Smith, B. A. (2015, April). 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* (pp. 797–806). Association for Computing Machinery.

Woodruff, A., Fox, S. E., Roussou-Schindler, S., & Warshaw, J. (2018, April). A qualitative exploration of perceptions of algorithmic fairness. In *Proceedings of the 2018 CHI conference on human factors in computing systems* (pp. 1–14). Association for Computing Machinery.

Xu, K., & Liao, T. (2020). Explicating cues: A typology for understanding emerging media technologies. *Journal of Computer-Mediated Communication*, 25(1), 32–43.

Youyou, W., Kosinski, M., & Stillwell, D. (2015). Computer-based personality judgments are more accurate than those made by humans. *Proceedings of the National Academy of Sciences of the United States of America*, 112(4), 1036–1040.

Zarsky, T. (2016). The trouble with algorithmic decisions: An analytic road map to examine efficiency and fairness in automated and opaque decision making. *Science, Technology, & Human Values*, 41(1), 118–132.

Zuboff, S. (2019). *Surveillance capitalism and the challenge of collective action*. SAGE.

**Author Biographies**

Tony Liao (PhD, Cornell University) is an assistant professor in the Department of Communication at the University of Cincinnati. His research interests focus on the development and effects of emerging technologies, specifically mobile media, augmented reality, and data algorithms.

Olivia Tyson is an undergraduate student in the Department of Communication at the University of Cincinnati. Her research interests focus on how people perceive social identities online.