Social Media Profiling Continues to Partake in the Development of Formalistic Self-Concepts. Social Media Users Think So, Too.

Severin Engelmann
Technical University of Munich
Garching, Germany
severin.engelmann@tum.de

Valentin Scheibe
Ludwig Maximilian University of Munich
Munich, Germany
v.scheibe@campus.lmu.de

Fiorella Battaglia
Università del Salento
Lecce, Italy
fiorella.battaglia@unisalento.it

Jens Grossklags
Technical University of Munich
Garching, Germany
jens.grossklags@in.tum.de

ABSTRACT

Social media platforms generate user profiles to recommend informational resources including targeted advertisements. The technical possibilities of user profiling methods go beyond the classification of individuals into types of potential customers. They enable the transformation of implicit identity claims of individuals into explicit declarations of identity. As such, a key ethical challenge of social media profiling is that it stands in contrast with people's ability to self-determine autonomously, a core principle of the right to informational self-determination.

In this research study, we take a step back and revisit theories of personal identity in philosophy that underline two constitutive meta-principles necessary for individuals to self-interpret autonomously: justification and control. That is, individuals have the ability to justify and control essential aspects of their self-concept. Returning to a philosophical basis for the value of self-determination serves as a reminder that user profiling is essentially normative in that it formalizes a person's self-concept within an algorithmic system. To understand whether social media users would want to justify and control social media's identity declarations, we conducted a vignette survey study (N = 368). First, participants indicate a strong preference for more transparency in social media identity declarations, a core requirement for the justification of a self-concept. Second, respondents state they would correct wrong identity declarations but show no clear motivation to manage them. Finally, our results illustrate that social media users acknowledge the narrative force of social media profiling but do not strongly believe in its capacity to shape their self-concept.

CCS CONCEPTS

• **Social and professional topics** → **User characteristics**; • **Security and privacy** → **Social aspects of security and privacy**.

KEYWORDS

Ethics of artificial intelligence, user profiling, personal identity, social media, autonomy.

ACM Reference Format:
Severin Engelmann, Valentin Scheibe, Fiorella Battaglia, and Jens Grossklags. 2022. Social Media Profiling Continues to Partake in the Development of Formalistic Self-Concepts. Social Media Users Think So, Too.. In *Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society (AIES'22), August 1–3, 2022, Oxford, United Kingdom.* ACM, New York, NY, USA, 15 pages. https://doi.org/10.1145/3514094.3534192

1 INTRODUCTION

Social media platforms enable advertisers to create and target user audiences based on the identification, processing, and analysis of several thousand user attributes such as likes, interests, beliefs, behaviors, relationships, moral convictions, and political leanings [3, 18, 50, 55, 65, 69]. User profiling techniques infer identity claims of users based on views and clicks, visual data such as images and videos, or the number and types of "followers" or "friends" [11, 18, 20, 21, 39, 45, 47, 70]. There is growing recognition in user profiling and user modeling communities that such profiling techniques create unique ethical challenges [3, 30, 63].

These challenges typically fall back on the inability of users to access, understand, and contest automatically-generated identity claims based on their personal data. Specifically, they arise from the restricted ability of social media users to exercise their right to informational self-determination, a central right of many privacy laws around the world. The right to informational self-determination rests on the fundamental idea that it is critical for individuals to freely and autonomously "self-determine" or "self-develop" [8, 36, 46, 49, 59]. The right to informational self-determination mandates that it is critical for individuals to be able to exercise control over their personal information. In the face of technologies that analyze the sentiment of users based on speech or visual data [18, 19, 57, 58] or that interpret data that users have shared unintentionally [2], the notion of individual control over personal data as a feasible mechanism for informational self-determination is, however, severely challenged.

In this paper, we offer a partly philosophical and a partly empirical account to address this problem field. From a philosophical perspective, we aim to make the following two contributions. First, we return to scholarship on the fundamental value of autonomous

AIES'22, August 1–3, 2022, Oxford, United Kingdom
© 2022 Copyright held by the owner/author(s).
This is the author's version of the work. It is posted here for your personal use. Not for redistribution. The definitive Version of Record was published in *Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society (AIES'22), August 1–3, 2022, Oxford, United Kingdom*, https://doi.org/10.1145/3514094.3534192.

self-determination offered by philosophical theories of personal identity. Philosophical theories of personal identity conceptualize necessary *procedural criteria* that enable an individual to form a self-concept. Personal identity is an essentially contested concept and, as such, inherently procedural—disputes on the concept's boundaries are essential to the concept itself [42, 54].[1] In contrast, when essentially contested concepts become subjected to digital formalism, they are fixated by definitions that work optimally only under the constraints of computability. The analysis of theories of personal identity can illustrate to us, perhaps again, the enormous power of social media user profiling in determining all procedural elements that exist between personal data and their analysis as declarations of identity: the power to create user profiles over time, the power to change or correct user profiles when needed, as well as the power to change the rules by which user profiles can be generated, changed, or corrected.

Second, the generation of digital representations of personal identity necessarily creates normative trade-offs. We present one normative trade-off by referring to what we call "model fitness." Here we ask whether the digital representation of an individual's self-concept *should* align as much as possible with how a person would self-determine in order to respect that person's autonomy. Social media platforms have the power to decide what types of data and what amounts of data are sufficient to justify an identity claim about a user. Social media platforms control "model fitness." We exemplify this phenomenon by referring to the literature on "window sliding" in learning tasks with concept drift adaptation [26, 38, 41, 71] and collaborative filtering [29].

We further take it that the power of social media profiling to make identity claims about billions of users is a strong argument in favor of usable transparency that allows users to view (understand their justification) and correct (exercise control over) such identity claims. Here, we engage in another trade-off: if people could view and correct identity claims of social media profiling, then such identity claims could influence a person's self-concept. Social media identity claims could undermine a person's autonomy to self-determine under conditions of transparency when people see, reflect on, and internalize "how a machine interprets" them. Transparency could empower social media identity claims rather than people's autonomy to self-determine.

Subsequently, we have conducted an empirical vignette study to understand how individuals (N = 368) evaluate social media's identity claims with regard to accuracy, transparency, and control. We find that people believe social media user profiling can make accurate judgments about essential aspects of their personal identity, but that they prefer privacy over accuracy. Moreover, people show a strong desire for transparency defined as the ability to view and understand declarations of identity by social media platforms. While people state that they want to compare whether such identity declarations align with their own self-concept, they believe that these do not influence their self-concept. Our study provides evidence that people assert that social media identity claims do not feed back into their own self-concept when they are made transparent and intelligible.

With this work, we seek to contribute to scholarship on the relation and interaction between humans and their algorithmically generated identity declarations. We provide a philosophical lens on the value of self-determination as the process to justify and control essential aspects of a person's self-concept. The conceptualization of autonomy through personal identity creates a firm foundation for determining the ethical challenges of social media user profiling. With a vignette survey study, we take a tangible step towards understanding how people actually evaluate algorithmic identity declarations by social media platforms.

Before we move on to the next section, we would like to offer a disclaimer: In this work, we do not claim that social media user profiling *generates* personal identity or suggest that the resulting profiles can be considered as equal to a person's self-concept. We do not engage in arguments that draw an ontological comparison between a user profile and the person behind it. In other words, we do not claim that social media user profiling leads to a user profile that *is* the personal identity of the individual. Rather, we observe that social media user profiling procedures possess a unique, technologically-afforded narrative force that computationally fixates the interpretative potential of a person's self-concept. This fixation creates ethical challenges when user profiling algorithms turn a person's personal data into declarations of identity that a person cannot view, cannot understand, and cannot contest.

2 SOCIAL MEDIA USER PROFILING IS FUNDAMENTALLY NORMATIVE

Our analysis considers user profiling procedures for *social media* advertisement. All major social media platforms offer a marketing page with an interface[2] where marketers can select desirable user attributes[3] for targeted advertising.

Previous work on social media profiling has summarized what kind of user attributes social media profiling generates. Such profiles consist of user inferences based on online data (e.g., user-generated content on the platform) as well as offline data (e.g., data integrated from data brokers) [3, 65]. User profiling for social media generates sophisticated representations of users based on demographic information including age or gender as well as information associated with user behaviors, preferences, and intentions [1, 12, 16, 27, 68]. Inferences are in part based on "explicit identity claims" (e.g., explicitly *stated* profession or sexual orientation) as well as on "implicit identity claims." Implicit identity claims are "given off" by an individual rather than consciously communicated [56, 62]. Implicit identity claims are inferences users communicate indirectly, for example, through their affiliations to certain individuals, social or institutional groups, preferences, and interests expressed in a non-specific manner. Explicit and implicit identity claims can comprise behaviors (e.g., clicks or views) and beliefs (expressions of interest, intentions, convictions, etc.) [3, 65]. Social media targeting tools offer marketers the option to select an audience (a group of users) based on whether they "possess" or do not "possess" a desirable attribute. Aimeur has provided a comprehensive list of the types of attributes (i.e., identity claims) analyzed for user profiling including

[1] Please note that this account focuses exclusively on Western approaches to philosophical theories of personal identity.

[2] See, for example, Meta audience insights or Instagram audience insights.

[3] We refer to such user attributes as "declarations of identity."

name, age, address, identity of friends, sexual orientation, political views, smoker yes/no, pregnancy/wedding, interests, credit score, home value, and others [2]. To understand the normative dimensions of user profiling on social media, the technological *instantiation* of a user's profile, for example as a feature vector [7], is not significant for this analysis. What is relevant is the algorithmic mapping function implemented to assign attributes to users based on their data. Any mapping process from user data to user inference digitally fixates the interpretative potential of an individual user. We refer to this process as the generation of a *formalistic self-concept*. By essentially determining this interpretative potential within an algorithmic frame, mapping functions become normative, for example, when they prioritize user data to constitute an attribute while failing to consider others.

In philosophy, a person's self-concept is procedural, contextual, and contestable [5, 10, 23]. Recent work in Science and Technology Studies has outlined that profiling socially contested concepts through mathematical formalism without accounting for their full meaning creates so-called abstraction and formalism traps [54]. Abstraction and formalization necessarily involve a process of imperfect translation: no model (or profile) is large enough to include all characteristics of an informational object. Similarly, in philosophy, no single theory of personal identity contains all constitutive principles that make up personhood. Indeed, it is the disagreement on fundamental conceptual features that creates the essential demarcations of a contested concept such as freedom, privacy, autonomy and so on [42]. In user profiling for social media advertisement, abstraction is constrained by two core conditions: First, by the purpose for which the object is profiled—here for commercial purposes (marketing)—and, second, by the mathematical constraints of computability. Regarding the latter, not all features of an object can be modeled by computational resources; for example, the phenomenological experience of human consciousness cannot— in principle—be captured by computational means.[4] Overall, philosophical theories of personal identity offer a useful conceptual framework to understand the normativity of generating formalistic self-concepts.

3 JUSTIFICATION AND CONTROL: TWO META-PRINCIPLES OF PERSONAL IDENTITY

In the following section, we detail how three influential theories of personal identity lay out procedural criteria that enable a person to form a self-concept autonomously.[5] Attributable to philosophical scholarship, such procedural requirements are subject to productive dispute. Yet, a body of philosophical scholarship on personal identity [22, 51–53, 60, 61] agrees on two constitutive meta-principles necessary for individuals to self-interpret autonomously: individuals have the ability to *justify* and *control* essential elements of their self-concept.[6] Some philosophers place the source of individuals' abilities to justify and control essential aspects of their self-concept in the individual only (e.g., [10, 37]); other theorists argue that social agents partake in the formation of a self-concept [51, 52, 60, 61].

3.1 Harry Frankfurt's second-order desires

In "Freedom of the Will and Concept of a Person," Harry Frankfurt developed a notion of personal identity grounded in the structure of human will [22]. Humans are capable of evaluating the desirability of their desires. A person also cares about the desirability of their desires. Frankfurt calls such desires "second-order desires" that are desires about desires or wants about wants. The object of a first-order desire is a state of affair, while a second-order desire's state of affair is a first-order desire. The desirability of our desires is ethically significant. For example, a person can want to want to eat in a certain way. Vegetarianism, an ethical principle, governs how a person acts on their first-order desire to eat. Frankfurt argues, "*only humans are capable of reflective self-evaluation manifested in the formation of second-order desires*" [22]. The essence of a person lies in will, however, a person needs to be able to "*become critically aware of their own will*" [22]. Individuals need critical reflection to evaluate which of their desires are desirable. Persons are autonomous in determining which desire they want to be moved by when acting. Repeated identification with a specific second-order desire enables us to truly care for something.

Frankfurt's theory of personal identity clearly presents a strong ideal of what it means to be a person. Individuals are required to engage in reflective justification of their second-order desires to fully qualify as persons. There is little room for ambiguous or even paradoxical desires that clearly constitute human experiences. Frankfurt's conception of personhood is an example of a theory from "within": his principles of personal identity are subjective and can even be criticized as "solipsism." External influences, cultural or social, appear to restrict rather than help strengthen individuals' ability to form a self-concept. Summarizing, Frankfurt's second-order desires stress the need for *justifying* one's self-concept, while the identification with a second-order desire underscores that persons can *control* what principles constitute their self-concept.

3.2 Charles Taylor's weak and strong evaluator

The philosopher Charles Taylor deliberately tries to avoid "solipsistic tendencies" and points to the importance of social interaction for the development of a self-concept. Taylor stresses the significance others have for our capacity to evaluate what we desire [60]. Many of our desires, wishes, hopes, attitudes, goals and so on develop only in dialogue with others. Taylor places personal identity between private and public spheres: Privately, a human being is a person because of their reflective self-evaluative capacities that require qualitative articulacy. Publicly, a person necessarily adopts such qualitative articulacy by interaction with other individuals.

[4] Theories on the phenomenological self by Dan Zahavi [67] develop a notion of personal identity that falls back on phenomenological experience.

[5] Personal identity conceptually differs from theories of *personality*. An account of personality is, for example, the prominent Big-Five (BFM) model of personality [32]. The BFM subscribes to personality theories that suggest personality to consist of context-consistent, quantitatively-assessable, enduring traits. In contrast, personal identity explains *how* individuals come to form a persistent self-concept. While such a self-concept may comprise a set of traits, it is the set of principles by which an individual's self-concept develops that is the focus of philosophical theories of personal identity.

[6] Other conceptualizations of hermeneutic personal identity also highlight—in some way or another—the importance of the two meta-principles of justification and control for a person's self-concept (see, for example, [5, 24, 64]). However, they motivate these principles with a different set of reasons.

Similar to Frankfurt's first-order desires and second-order desires, Taylor distinguishes between so-called "weak" and "strong evaluators"[7]. A weak evaluator simply deliberates different options on the basis of their convenience: their goal is to get the most overall satisfaction. Such an evaluator does not reflect on the qualitative aspects of their choices. Non-qualitative evaluation leads to the selection of a desired object or action because "*of its contingent incompatibility with a more desired alternative*" [61]. A weak evaluator chooses something merely on circumstantial grounds. Their deliberation does not exceed a mere desirability calculation for choices to provide some satisfaction. Taylor claims that persons can evaluate what they are and shape whatever they wish to be on this basis. Different from Frankfurt, however, the freedom to self-interpret takes place between private and social spheres. This freedom (i.e., control) to self-define by evaluation (i.e, by justification) means that persons can be made *responsible* for their self-concept [61].

3.3 Maya Schechtman's narrative self-constitution view

The philosopher Marya Schechtman asserts that an autonomous person has the capacity to psychologically organize a stream of events into a culturally accepted form of a narrative "*by which we will come to think of ourselves as persisting individuals with a single life story*" [52]. The elements of a narrative that a person can articulate constitute the person to a higher degree than those elements that a person cannot articulate.

An individual compares, organizes, and relates experiences by culturally-determined standards. It follows that no time-slice—any momentary event that an individual experiences—is in any way definitive for a person's identity. Only when interpreted in the context of the narrative is such a time-slice a meaningful element of a person.[8] Telling a story is only one element of a person's narrative. Individuals form a narrative, but they also enact it and subsequently criticize it: they are not only the authors of their narrative but their protagonists and critics, too. As an author, a person tries to understand the meaning events have by integrating them into their continuous narrative. A person is the critic of their narrative when they come to reflect, evaluate, and criticize the actions they have carried out. While the order in which these steps take place is certainly dynamic, it demonstrates that a person plays different roles within their own narrative—they are not simply describing what they have experienced as a commentator or storyteller in the literal meaning of the term. For Schechtman, a person's narrative is actively *negotiated* between subjective and objective accounts. A person may have their own interpretation of a certain event; however, their identity will be undermined if claims reach a level of incomprehensibility for other people. A person's choices and actions must "*flow intelligibly from (their) intentions, motives, passions, and purposes...*" [52]. Without our narrative context, other individuals cannot make sense of our choices and actions. The narrative view gives individuals freedom to shape (i.e., control) who they wish to be, re-interpret their past and anticipate their future self-concept (i.e., justification). A person's social environment holds a person accountable for the narrative they articulate.

Summary of philosophical theories of personal identity: While differences exist between the theories by Frankfurt, Taylor, and Schechtman, two meta-principles can be discerned: justification and control. First, a self-concept develops through *reflective justification*. Individuals become persons when they justify their self-concept—through reflective capabilities and in a narrative that is negotiated between subjective and objective accounts. Second, individuals can exert some control over their self-concept. While the theories disagree over the degree of control individuals have in forming an understanding of themselves, fundamentally, they all suggest that personhood is grounded in an individual's autonomy to determine essential aspects of their hermeneutic identity. It is for this reason that persons can justifiably be held responsible for their own identity.

4 TWO NORMATIVE TRADE-OFFS IN USER PROFILING FOR SOCIAL MEDIA MARKETING

We argue that social media profiling generates digitally formalized identity claims of a person by mechanisms that do not sufficiently allow for justification and control. In the following, we discuss two normative trade-offs that result from the inherent normativity of social media user profiling as discussed in Section 2.

4.1 Normative trade-off 1: The privacy versus model fit trade-off

4.1.1 Concept drift challenges. One normative judgment user profiling is necessarily required to make is to determine when enough data (or evidence) has been collected and analyzed to justify the inference of a person's attribute (i.e., an identity declaration). It is a normative undertaking to decide when the amount of personal data is sufficient to ensure proportionality between the user input and the attribute inference. Is the inference proportional to a single activity or expression of belief? Or is its proportionality dependent on multiple consecutive expressions of the belief? Resolving such questions, user profiling necessarily excludes user input from being considered for drawing user inferences. Schechtman asserts that individuals have the capacity to attribute meaning to a selection of experiences that become part of their own unique narrative. However, it is the narrative that is self-constituting, not the single experience. It follows that no time-slice—any momentary event that an individual experiences—is in any way definitive for a person's identity. Such a time-slice is only a descriptive and meaningful element of a person when interpreted in the *context* of the narrative.

Schechtman's concept of a "time-slice" can be compared to the concept of "window sliding" used in learning tasks with concept drift adaption [26, 38, 41, 71]. Concept drift techniques are deployed to gain knowledge from data stream *changes*. Drifts or changes in a data stream can be either sudden or gradual. The former could be a sudden new interest in a new subject, while the latter could be a growing interest in moving to another country. In user profiling, concept drift belongs to a class of challenges called dynamicity problems [48, 66]. Recommender systems apply dynamic user profiles to offer more value to the user, who sees informational resources

[7] Arguably, a person that chooses merely on the basis of Frankfurt's first-order desires corresponds to Taylor's weak evaluator.

[8] "*Whether or not a particular action, experience, or characteristic counts as mine is a question of whether or not it is included in my self-narrative*" [25].

they have only recently become interested in, and to the advertiser that can bid for audiences with the most up-to-date profile.

Machine learning (ML) classifiers are able to respond to concept drift—gradual, sudden, or reoccurring changes often in multiple data streams—without "neglecting" the outdated data [71]. For example, sliding windows of fixed and variable sizes of training data are used to build an updated model [26]. Since both fixed and variable windows are definite in their size, some old data will necessarily be "forgotten." What criteria determine which data are to be forgotten and which ones are to be considered in creating an updated profile of a person? The promise of targeted advertisement rests on the belief that more recent user data corresponds to a more accurate profile of the user. However, model fit, a continuously updated model of a user's profile, requires a potentially uninterrupted flow of user data, raising privacy concerns [13]. The more time-slices are created, the more accurate the representation of the user, but the more user data is needed.

4.1.2 Lookalikes through Neighborhood-Based Collaborative Filtering. Collaborative filtering (CF) is one of the most widely applied user modeling techniques in many recommender system. For example, as a user profiling technique, *k*-nearest neighbor relies on the assumption of similarity between individuals [29]. Similar profiles presumably react similarly to certain informational items. The advantage of CF is that one only requires a model of one of the two—users or items—to model the other. Consequently, CF uses items to model users and users to model items. The more users evaluate informational resources, the more they help the system for its predictive analysis of other users. Social media (as well as search engines) offer their customers so-called "Lookalike Audiences."[9] With many marketers, Lookalikes are popular since they can use their well-known customer base to target "similar" but potentially new customers. Lookalikes are less privacy-invasive because they use data that is already available to make inferences about a user. Taylor's and Frankfurt's concept of a person, however, stresses the ability of persons to decide what is *desirable for them*. *k*NN-based CF and Lookalikes work in the opposite way. They determine the desirability of one's desires as equal or at least similar to the desirability of other, already "known" individuals' desires, to use Frankfurt's nomenclature.

4.2 Normative trade-off 2: The transparency versus autonomy trade-off

A key question is if people would actually care about model fit—an accurate representation of their *formalistic* data narrative. Perhaps individuals do, after all, live in the best of all possible worlds: they draw enormous benefits from using social media and do not worry about how their data is mapped to a spectrum of attribute inferences. One way forward would be to enable individuals to understand and correct inferences they do not agree with. Here, another normative complication emerges. A person could gain autonomy from having access to their social media's identity declarations. However, these identity declarations could in turn influence a person's self-concept.

Should individuals get access in order to understand and contest their "data narrative"? Providing explanations on "how the systems works" has shown to increase users' trust in many different recommender systems [9, 14, 17, 44]. Usable transparency allows users to tell the system when an inference is presumptuous (or even wrong). For example, a system could show users those identity declarations that have been sold to marketers or that were based on implicit identity claims. However, simply revealing—at least in part—the content behind user profiles could support internalization and conformation to the proposed inferences. Perhaps individuals would welcome such a degree of transparency as a mechanism to "offload" the psychological work necessary to attribute meaning to certain life events posted online [51–53]. Making inferences transparent to the individual means recognizing their semantic power in shaping who individuals are and who they can become. This second normative trade-off arises from the question of whether the autonomy gained from being able to understand such recommended inferences outweighs a potential loss of autonomy when they become part of a person's self-concept. This could mean that, today, a person, their social network (offline and online), *and* social media profiling identity declarations together participate in creating a person's self-concept.

The effect on individuals' self-concept could be enhanced if social media user profiling generates specific identity declarations repeatedly or even permanently. According to Frankfurt's theory of personal identity, a person attempts to form a self-concept that stems from their care for what they desire. Frankfurt recognizes that one can only care about something if it is for extended periods of time. Desires typically last for moments only: if one cared about something for only a moment one could not be distinguished from a person that acted out of impulse. How would users perceive such recommended attribute inferences? Perhaps with little skepticism, since they would acknowledge the algorithmic output as an objective and truthful interpretation of their wishes, wants, and desires?

5 METHODS AND EXPERIMENTAL PROCEDURE: VIGNETTE STUDY

To address the key questions arising from both normative trade-offs, we conducted a vignette study that asked respondents a) whether they believed social media profiling could accurately infer elements of their self-concept, b) whether they considered accuracy of these identity declarations to be desirable, c) whether they had motivation to view and correct identity declarations, and d) whether they believed that social media identity declarations would influence their self-concept if they were made transparent to them. The goal of the vignette study was to take a tangible step towards understanding whether social media users preferred accuracy of social media identity declarations over privacy (trade-off 1) and whether they believed that social media identity declarations would influence their self-concept (trade-off 2). Vignette studies have been extensively used in human computer interaction, psychology, and experimental philosophy to elicit participants' explicit ethical judgments in various hypothetical scenarios [4, 6, 15, 28, 31, 33, 34, 40, 43]. Moreover, with our vignette survey study, we follow calls for more experimentally-informed AI ethics [35].

Our study was a within-subject design, we presented each respondent with the same hypothetical vignette scenario. First, the

[9] See, for example: https://www.facebook.com/business/help/164749007013531 accessed May 30, 2022.

vignette asked respondents to imagine that they are active users on a social media platform (see the hypothetical vignette scenario in Appendix A.1). As an active user, each respondent was told that they regularly engage in typical actions on the social media platform. Participants read that they publish postings, share postings by other users, and react to other users' postings. Second, the vignette introduced examples of data types each respondent shares with the social media platform (gender, location, relationship status, social contacts, content viewed, content clicked, etc.). Respondents were told that the social media platform uses algorithms to draw conclusions about them based on the data they share in order to show them more suitable content and advertisements. Third, the vignette elaborated on the types of conclusions (i.e., identity declarations) that the social media platforms draws about them. The vignette explained that the platform collects data that users actively share to draw conclusions about them. For example: *"...when you provide your real birthday, the platform uses this information to show you content that it takes to be suitable for your age group."* Respondents were also told that the platform draws conclusions about users based on data that users may not be aware that they are sharing. For example, *"...since you share your location data, the platform tries to conclude where you work and live. As another example, the platform also tries to conclude what hobbies you have based on your friends' activities."* Respondents were further told that, using their data, the social media platform attempts to conclude their interests, their political orientation, their religious beliefs, and aspects of their personality (among others). Lastly, we asked respondents to imagine that the platform "combines and stores" all conclusions about them in a so-called "user data profile" (UDP). The vignette explained to users that the social media platform uses the content of their UDP to recommend relevant information and advertisements. The hypothetical vignette scenario ended by telling respondents that the social media platform generates all of its revenues from personalized advertisement. We included two attention checks in the vignette. All participants were active social media users.

After respondents had read the vignette and passed the attention checks, they rated questions using a 7-point Likert scale. Questions were divided into 5 categories and shown to respondents in random order within these categories. The first two categories of questions asked respondents whether they believed social media platforms could make accurate judgments about them and whether the social media platform *should* make accurate judgments about them. We defined accuracy as a) general judgments, b) specific judgments, and c) temporal judgments. The third and fourth set of questions asked whether respondents desired to view and understand social media judgments about them and whether they would change incorrect judgments. Questions on respondents' preference for transparency included a) data collection & use, b) preference for understanding conclusions of the social media platform, and c) preference for transparency of their UDP (i.e., all identity declarations). Finally, a fifth set of questions asked respondents whether social media judgments would have an influence on their self-concept given that respondents could view their UDP. We defined "influence" as respondents' willingness to a) compare elements of their UDP with their self-concept, b) their willingness to reevaluate their self-concept in light of the identity declarations in their UDP, and c) their willingness to integrate elements of their UDP into their self-concept that they would not have associated with their self-concept. All questions are listed in Appendix A.3.

We recruited participants with Prolific. Based on pretests, we set the expected completion time at 20 minutes, with a payout of USD 3.75 (above US minimum wage of 2021). Data collection started on July 26, 2021 and ended on August 8, 2021. We recruited 458 respondents from the United States user base. 59 submissions were excluded for failing one of two attention checks, 10 for duplicate submissions, 9 for an unusually short response time, and 11 for being invalid (e.g., no prolific ID). This resulted in a final sample of 368 respondents (see demographics in the Appendix A.4). The mean time of completion was 15.3 minutes.

Our home institution does not require an ethics approval for questionnaire-based online studies. When conducting the study and analyzing the data, we followed standard practices for ethical research: presenting detailed study procedures, obtaining consent, not collecting identifiable information or device data, and using a survey service[10] that guaranteed compliance with the European Union's General Data Protection Regulation. The study did not include any deceptive practices. Subjects could drop out of the study at any point. All data were fully anonymized, and the privacy of all subjects was maintained at all times during the study.

6 RESULTS

Respondents' beliefs on the ability of social media platforms to make accurate judgments about them (Fig. 1a). A majority of respondents believed that social media algorithms could make accurate and correct judgments about them *in general* (78.2%). While 66.2% of respondents were convinced that social media algorithms could correctly judge "what is valuable to them," just over half of respondents said that social media algorithms can accurately reflect who they are (51.1%). Most respondents believed that their UDP was unique in comparison to other social media users (72.6%). However, only a minority of respondents said that family and close friends would be able to identify them by their UDP (45.5%).

Respondents' beliefs on the ability of social media platforms to make accurate judgments about them on specific attributes (Fig. 1b). Respondents believed that social media algorithms can accurately infer their interests (89.9%), their past (81.3%) and future purchasing behaviors (64.5%), as well as their location (77.4%). Just over half of those surveyed stated that social media algorithms could accurately conclude who they meet (54.8%).

Respondents also said that social media algorithms are able to accurately conclude their political stance (80.5%) and, albeit with less agreement, their religious beliefs (59.5%). Most respondents agreed that social media algorithms can correctly infer their attitudes towards the COVID-19 vaccine (77.9%), climate change (74.6%), and immigration (64.8%). However, respondents did not think that social media profiling was able to differentiate between their private and social self both online (35.5%) and offline (30.7%).

Respondents' beliefs on the ability of social media platforms to make accurate temporal judgments about them (Fig. 1c). Respondents believed that social media algorithms are able to keep their UDP up to date (71.4%). Respondents stated that

[10] SoSci Survey: https://www.soscisurvey.de/

Figure 1: (a) Respondents believe social media platforms (SMP) can make accurate judgements about them. UDP—user data profile. (b) Respondents believe social media (SM) algorithms are able to accurately infer a variety of attributes including their interests, purchases, location, political stance, or religious beliefs. Respondents do not believe SMP is able to distinguish who they are in private vs. who they are in social contexts. (c) Respondents believe SMP is able to keep their UDP up to date, but that their UDP does not tell an accurate story of their life. Note for all figures: results for "strongly agree" and "agree" are shown as "agree," results for "strongly disagree" and "disagree" are shown as "disagree."

their UDP from a month ago still included accurate conclusions (69.1%). However, just over half of respondents thought that their UDP from a year ago was still accurate (51.4%). A majority of respondents said that the social media platform would be able to conclude whether they had changed as a person after several years of being a user (68.9%). In contrast, only a minority of respondents believed that their *entire UDP* would tell an accurate story of their life since they started using the platform (37.9%).

Respondents' beliefs on the normativity of accurate social media judgments (Fig. 2). Most respondents stated that they wanted social media platform operators to ensure that their UDP was accurate (72.4%). Just more than half of respondents wanted social media operators to invest extra resources to make sure their UDP was accurate (56.9%). However, only a minority of 28.8% of respondents were in favor of trading their personal data for the creation of their UDP. Importantly, respondents did not want to

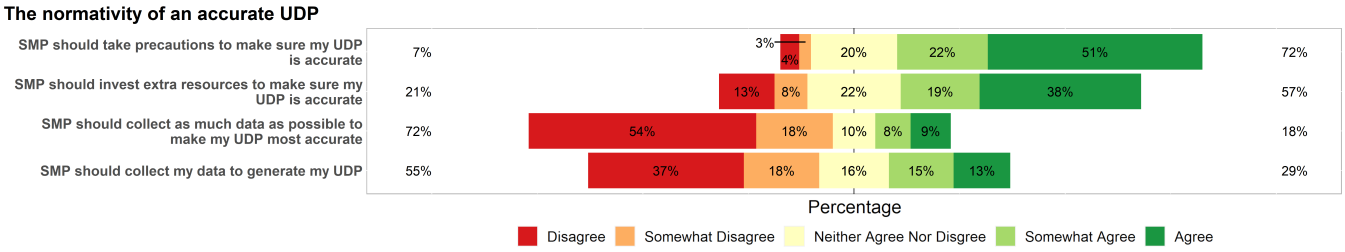

Figure 2: Respondents prefer an accurate UDP but not at the expense of their privacy.

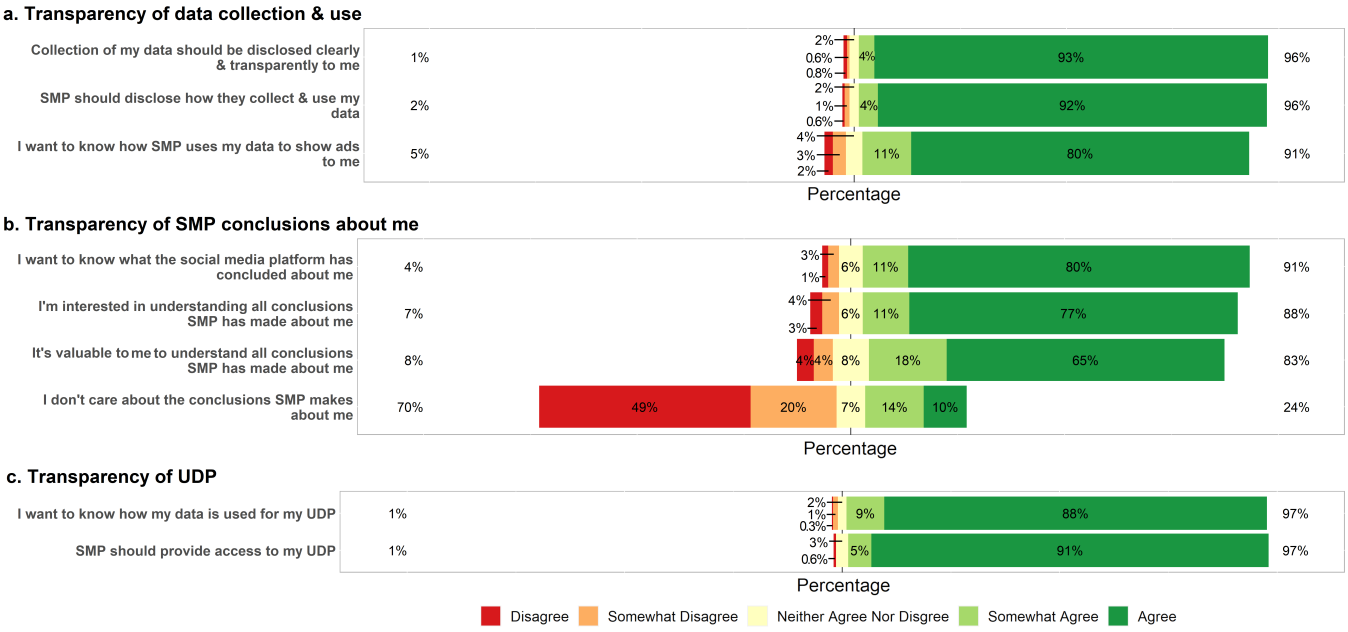

Figure 3: Respondents show great preference for transparency of (a) personal data collection & use, (b) conclusions SMP has made about them, and (c) of their UDP.

trade their personal data for an accurate UDP: only 17.9% agreed that the social media platform should collect as much personal data as possible to ensure that their UDP was as accurate as possible.

Respondents' preference for transparency of data collection & use (Fig. 3a). Respondents expressed their desire for transparency of personal data collection on social media, transparency of conclusions the social media platform made about them based on their data, and transparency of their UDP. Regarding data collection, most respondents stated that procedures of data collection should be disclosed clearly and transparently to them (96.4%) and that the social media platform should disclose how they collected and used their personal data in general (96.1%) and for showing advertisements (91.1%).

Respondents' preference for transparency of conclusions (Fig. 3b & c). Similarly, respondents showed a strong preference to understand what the social media platform has concluded about them (90.1%). Of the respondents, 87.7% stated that they were interested in understanding all conclusions the social media platform had made about them and 83.2% believed that such an understanding would be valuable to them. Only 24% of respondents stated that they do not care about conclusions the social media platform draws about them. Finally, similarly large majorities of respondents expressed their desire to understand how their personal data was used to create their UDP (96.7%). Of the respondents, 96.6% said that they wanted access to their UDP in general (Fig. 3c).

Respondents' preference for control over their UDP (Fig. 4). While respondents showed a clear preference for transparency, their desire to control (i.e., change or otherwise influence) their UDP was mixed. A majority stated that the social media platform should allow them to correct errors in the their UDP (90.2%). However, only a small majority said they would be motivated to change wrong conclusions in their UDP (60.2%). When we asked whether correcting and maintaining their UDP would be "too tedious," respondents showed no clear preference (agree: 37.8% vs. disagree: 45.4%, neither: 16.8%). Approximately half of respondents (57.3%) believed they

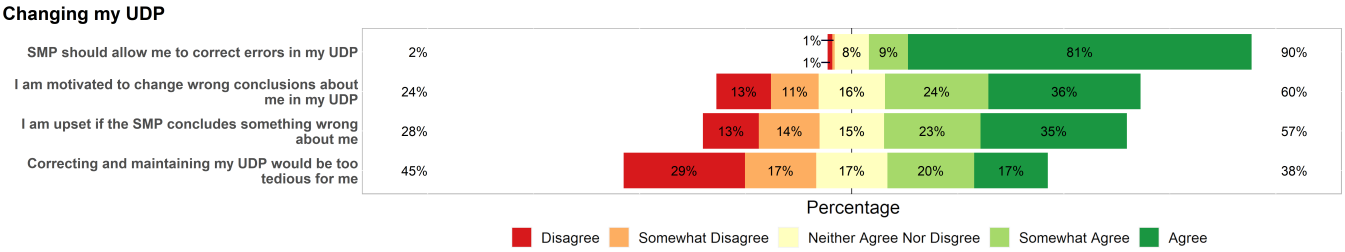

Figure 4: Respondents state that the SMP should allow them to correct errors in their UDP but provide no clear preference on whether they would be willing to correct and maintain their UDP.

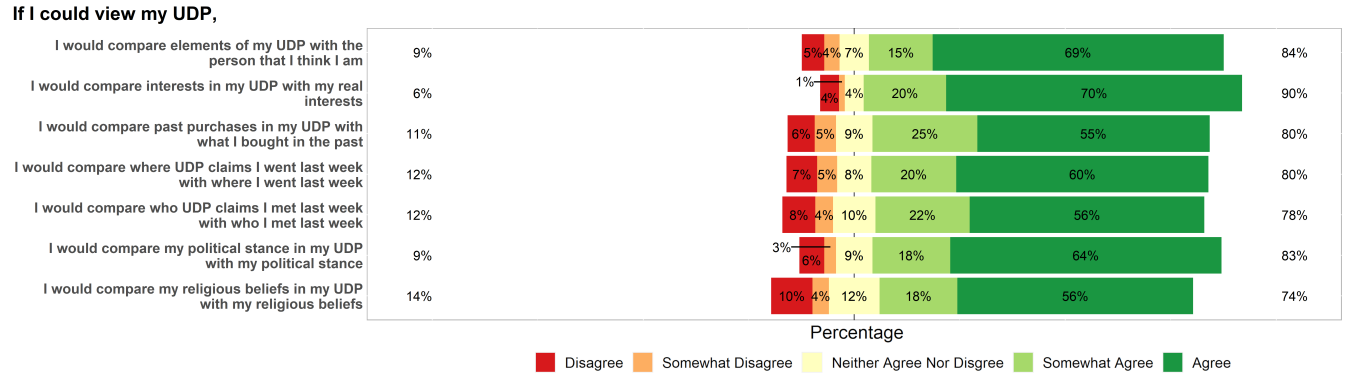

Figure 5: Provided their UDP was transparent, respondents would compare elements of their UDP with a range of personal attributes.

would be upset if the social media platform concluded something about them that they thought was incorrect.

Respondents' beliefs on the influence of the UDP on their self-concept (comparison UDP vs. self-concept, Fig. 5). Provided they had access to their UDP, the majority of respondents maintained that they would compare elements of their UDP with the person they thought they were (84.1%). Most respondents said they would compare interests in their UDP with their real interests (89.7%). Respondents further stated they would compare past purchases (79.9%), past locations (79.9%, "last week"), and past social meetings (77.9%, "last week") with those in their UDP. Among the respondents, 82.7% would compare their political stance with the one registered in their UDP and 74.3% of respondents would compare their religious beliefs with those in their UDP.

Respondents' beliefs on the influence of the UDP on their self-concept (reevaluation of self-concept, Fig. 6). Only a minority of respondents believed that viewing their user data profile would result in a reevaluation of their self-concept (agree: 21.5%). Few respondents stated that they would reevaluate their interests (agree: 17.3%), their future purchases (agree: 26.8%), their political stance (agree: 11.5%) or their religious beliefs (agree: 9.22%) after viewing their UDP.

Respondents' beliefs on the influence of the UDP on their self-concept (meaning of unaware identity declarations in the UDP, Appendix Fig. 7). Respondents were undecided whether social media conclusions were meaningful to them (agree: 47.3% vs. disagree: 35.6%). A small majority of respondents disagreed that conclusions about them in their UDP—that they did not know about—would be meaningful to them (disagree: 53.8%). A small majority of respondents also objected to statements saying conclusions about their political stance (disagree: 55.0%) or religious beliefs (disagree: 59.9%) in their UDP—that they did not know about—would be meaningful to them. Finally, we asked respondents whether their UDP would be a source of inspiration when looking for a *new* interest. Only 41.1% of respondents said that they would look into their UDP for suggestions on new interests. Likewise, respondents believed that when they saw an interest in their UDP that they would not have believed to be their interest, then this "recommended" interest would not become a new interest for them (agree: 36.6%). An even smaller minority of respondents said that predicted purchases in their UDP would influence actual future purchases (agree: 34.6%).

7 DISCUSSION OF RESULTS AND CONCLUDING REMARKS

In this work, we argued that the computability of digital representations of personal identity creates normative trade-offs when social media profiling generates identity claims that work only under the constraints of computability and that people cannot understand, view, or contest. Consequently, one of the key ethical challenges of social media profiling is that it stands in contrast with people's ability to self-determine freely and autonomously. To illustrate the

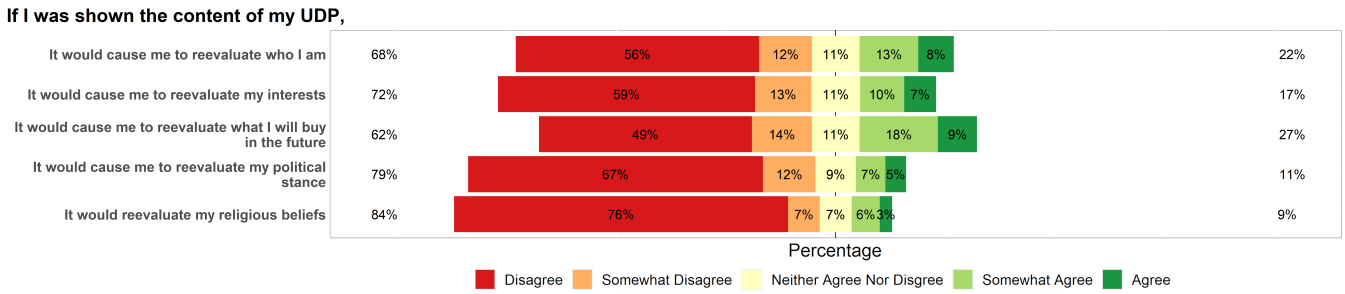

Figure 6: Respondents strongly believe that viewing the content of their UDP would not cause them to reevaluate elements of their self-concept.

inherently procedural nature of autonomous self-determination, we revisited theories of personal identity in philosophy that underline two constitutive meta-principles: justification and control. That is, individuals have the ability to *justify* and *control* essential elements of their self-concept. The return to the philosophical basis for the value of self-determination serves as a reminder that social media profiling represents an inherently normative formalization process of a person's self-concept. Within the interpretative space between data and declaration, social media platforms determine the meaning of views, clicks, posts, and social relationships without offering usable means for understanding or correcting essential parts of this process. As such, social media identity declarations are radically different from the procedural criteria laid out by theories of personal identity in philosophy.

Taking a step toward understanding how "ordinary" social media users view social media identity declarations, we conducted a vignette survey study. We found that people believe that social media platforms can make a variety of accurate judgements about them but that they cannot represent their entire self-concept. For example, respondents thought that social media profiling is able to accurately infer whether they have changed as a person over time, but that it cannot tell an accurate story of their life since signing up to the platform. Thus, respondents defined limits for the ability of social media identity declarations to represent certain aspects of their self-concept. Interestingly, respondents did claim that their own user data profile (UDP) was unique and that other users had a different UDP.

Respondents showed a strong preference for more transparency and stated that they would compare their own self-concept with a variety of social media identity declarations. However, the respondents in our study did not believe that social media identity declarations would be meaningful to them. Respondents also stated they would correct wrong identity declarations but showed no clear motivation to manage them. Taken together, we believe that it is reasonable to assume that social media users have at least some motivation to control essential aspects of their social media identity declarations. Providing such identity controls does present technological as well as design challenges for social media platform operators. However, social media platforms go to great lengths to offer advertisers usable controls to specify which user attributes exactly they wish to include in their custom audiences. In providing usable justification and control, social media platforms give priority to advertisers determining detailed custom audiences for targeted advertisement over giving users the possibility to understand, control, and rectify potential inaccuracies in their user profiles.

Finally, respondents did not believe that social media identity declarations would *influence* their self-concept. Respondents stated that previously unknown identity declarations would be unlikely to become part of their self-concept and they strongly objected that viewing social media identity declarations would cause them to reevaluate their self-concept. Future studies should try to understand whether people's self-concept is resilient to social media identity declarations as participants stated in our study. Perhaps people are overconfident in the immunity of their self-concept against social media declarations? Also, a majority of respondents expressed the desire to compare components of their UDP with their self-concept. Considering our results, we take it that people are, at least, curious to understand how social media platforms interpret them based on their personal information. They acknowledge the narrative force of social media profiling but do not strongly believe in its capacity to shape their self-concept. We encourage future studies to explore whether our findings extend to social media users in other cultures.

To conclude, we have focused on the *process* by which social media generate identity declarations based on personal information through user profiling. In comparison to the large corpus of studies that have focused on the *consequences* of user profiling (e.g., filter bubbles, misinformation), philosophical accounts on the procedural aspects of social media user profiling remain scarce. While our vignette study produces an initial understanding of the relationship between social media users and their identity declarations, we expect that this account provides ample opportunity for follow-up studies on the ethical challenges of social media profiling. Social media will continue to exercise its power to partake in the formation and development of formalistic self-concepts. We provide evidence that social media users think so, too.

ACKNOWLEDGMENTS

We are grateful to the anonymous reviewers for their constructive feedback and comments. This research was supported by a Volkswagen Foundation Planning Grant. The Volkswagen Foundation had no role in any part of the research or the decision to submit the manuscript for publication. The authors declare no competing or financial interests.

REFERENCES

[1] Ahmad Abdel-Hafez and Yue Xu. 2013. A survey of user modelling in social media websites. *Computer and Information Science* 6, 4 (2013), 59–71. https://doi.org/10.5539/cis.v6n4p59

[2] Esma Aïmeur. 2018. Personalisation and privacy issues in the age of exposure. In *Conference on User Modeling, Adaptation and Personalization (UMAP)*. 375–376. https://doi.org/10.1145/3209219.3209271

[3] Athanasios Andreou, Giridhari Venkatadri, Oana Goga, Krishna Gummadi, Patrick Loiseau, and Alan Mislove. 2018. Investigating ad transparency mechanisms in social media: A case study of Facebook's explanations. In *Network and Distributed System Security Symposium (NDSS)*. 1–15. https://www.ndss-symposium.org/wp-content/uploads/2018/02/ndss2018_10-1_Andreou_paper.pdf

[4] Kwame Anthony Appiah. 2008. *Experiments in ethics.* Harvard University Press.

[5] Kwame Anthony Appiah. 2010. *The ethics of identity.* Princeton University Press.

[6] Christiane Atzmüller and Peter M. Steiner. 2010. Experimental vignette studies in survey research. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences* 6, 3 (2010), 128–138. https://doi.org/10.1027/1614-2241/a000014

[7] Joseph Blass. 2019. Algorithmic advertising discrimination. *Northwestern University Law Review* 114, 2 (2019), 415–467. https://scholarlycommons.law.northwestern.edu/nulr/vol114/iss2/3/

[8] Edward J. Bloustein. 1964. Privacy as an aspect of human dignity: An answer to Dean Prosser. *New York University Law Review* 39, 6 (1964), 962–1007. https://heinonline.org/HOL/LandingPage?handle=hein.journals/nylr39&div=71

[9] Jens Brunk, Jana Mattern, and Dennis M. Riehle. 2019. Effect of transparency and trust on acceptance of automatic online comment moderation systems. In *2019 IEEE 21st Conference on Business Informatics (CBI)*. 429–435. https://doi.org/10.1109/CBI.2019.00056

[10] Sarah Buss and Lee Overton (Eds.). 2002. *Contours of agency: Essays on themes from Harry Frankfurt.* MIT Press.

[11] Buru Chang, Yonggyu Park, Donghyeon Park, Seongsoon Kim, and Jaewoo Kang. 2018. Content-aware hierarchical point-of-interest embedding model for successive POI recommendation. In *International Joint Conferences on Artificial Intelligence (IJCAI)*. 3301–3307. https://www.ijcai.org/proceedings/2018/0458.pdf

[12] Jinpeng Chen, Yu Liu, and Ming Zou. 2016. Home location profiling for users in social media. *Information & Management* 53, 1 (2016), 135–143. https://doi.org/10.1016/j.im.2015.09.008

[13] Michela Chessa, Jens Grossklags, and Patrick Loiseau. 2015. A game-theoretic study on non-monetary incentives in data analytics projects with privacy implications. In *IEEE 28th Computer Security Foundations Symposium (CSF)*. 90–104. https://doi.org/10.1109/CSF.2015.14

[14] Jong Kyu Choi and Yong Gu Ji. 2015. Investigating the importance of trust on adopting an autonomous vehicle. *International Journal of Human-Computer Interaction* 31, 10 (2015), 692–702. https://doi.org/10.1080/10447318.2015.1070549

[15] Cory J. Clark, Jamie B. Luguri, Peter H. Ditto, Joshua Knobe, Azim F. Shariff, and Roy F. Baumeister. 2014. Free to punish: A motivated account of free will belief. *Journal of Personality and Social Psychology* 106, 4 (2014), 501–513. https://doi.org/10.1037/a0035880

[16] Christopher Ifeanyi Eke, Azah Anir Norman, Liyana Shuib, and Henry Friday Nweke. 2019. A survey of user profiling: State-of-the-art, challenges, and solutions. *IEEE Access* 7 (2019), 144907–144924. https://doi.org/10.1109/ACCESS.2019.2944243

[17] Nadia El Bekri, Jasmin Kling, and Marco F. Huber. 2020. A study on trust in black box models and post-hoc explanations. In *14th International Conference on Soft Computing Models in Industrial and Environmental Applications (SOCO)*, Francisco Martínez Alvarez (Ed.). Advances in Intelligent Systems and Computing, Vol. 950. Springer, 35–46. https://doi.org/10.1007/978-3-030-20055-8_4

[18] Severin Engelmann and Jens Grossklags. 2019. Setting the Stage: Towards Principles for Reasonable Image Inferences. In *Adjunct Publication of the 27th Conference on User Modeling, Adaptation and Personalization (UMAP)*. 301–307. https://doi.org/10.1145/3314183.3323846

[19] Severin Engelmann, Chiara Ullstein, Orestis Papakyriakopoulos, and Jens Grossklags. 2022. What People Think AI Should Infer From Faces. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*. https://doi.org/10.1145/3531146.3533080

[20] Golnoosh Farnadi, Jie Tang, Martine De Cock, and Marie-Francine Moens. 2018. User profiling through deep multimodal fusion. In *Eleventh ACM International Conference on Web Search and Data Mining (WSDM)*. 171–179. https://doi.org/10.1145/3159652.3159691

[21] Bruce Ferwerda, Markus Schedl, and Marko Tkalcic. 2015. Predicting personality traits with Instagram pictures. In *Workshop on Emotions and Personality in Personalized Systems*. 7–10. https://doi.org/10.1145/2809643.2809644

[22] Harry G. Frankfurt. 1971. Freedom of the will and the concept of a person. *The Journal of Philosophy* 68, 1 (1971), 5–20. https://doi.org/10.2307/2024717

[23] Shaun Gallagher. 2000. Philosophical conceptions of the self: Implications for cognitive science. *Trends in Cognitive Sciences* 4, 1 (2000), 14–21. https://doi.org/10.1016/S1364-6613(99)01417-5

[24] Shaun Gallagher (Ed.). 2011. *The Oxford handbook of the self.* Oxford University Press.

[25] Shaun Gallagher and Dan Zahavi. 2020. *The phenomenological mind.* Routledge. https://doi.org/10.4324/9780429319792

[26] João Gama, Raquel Sebastião, and Pedro Pereira Rodrigues. 2013. On evaluating stream learning algorithms. *Machine Learning* 90, 3 (2013), 317–346. https://doi.org/10.1007/s10994-012-5320-9

[27] Susan Gauch, Mirco Speretta, Aravind Chandramouli, and Alessandro Micarelli. 2007. User profiles for personalized information access. In *The Adaptive Web*, Peter Brusilovsky, Alfred Kobsa, and Wolfgang Nejdl (Eds.). Springer, 54–89. https://doi.org/10.1007/978-3-540-72079-9_2

[28] Armin Granulo, Christoph Fuchs, and Stefano Puntoni. 2019. Psychological reactions to human versus robotic job replacement. *Nature Human Behaviour* 3, 10 (2019), 1062–1069. https://doi.org/10.1038/s41562-019-0670-y

[29] Jonathan Herlocker, Joseph Konstan, Al Borchers, and John Riedl. 1999. An algorithmic framework for performing collaborative filtering. In *22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. 230–237. https://doi.org/10.1145/312624.312682

[30] Mireille Hildebrandt. 2019. Privacy as protection of the incomputable self: From agnostic to agonistic machine learning. *Theoretical Inquiries in Law* 20, 1 (2019), 83–121. https://doi.org/10.1515/til-2019-0004

[31] Michael R. Hyman and Susan D. Steiner. 1996. The vignette method in business ethics research: Current uses, limitations, and recommendations. In *Proceedings of the Annual Meeting of the Southern Marketing Association*. 261–265.

[32] Oliver John, Laura Naumann, and Christopher Soto. 2008. Paradigm shift to the integrative Big Five trait taxonomy: History, measurement, and conceptual issues. In *Handbook of Personality: Theory and Research* (3rd ed.), Oliver P. John, Richard W. Robins, and Lawrence A. Pervin (Eds.). The Guilford Press, 114–158.

[33] Joshua Knobe. 2003. Intentional action and side effects in ordinary language. *Analysis* 63, 3 (2003), 190–194.

[34] Joshua Knobe and Shaun Nichols. 2017. Experimental philosophy. In *The Stanford Encyclopedia of Philosophy* (Winter 2017 ed.), Edward N. Zalta (Ed.). Metaphysics Research Lab, Stanford University.

[35] Steven R. Kraaijeveld. 2021. Experimental philosophy of technology. *Philosophy & Technology* 34, 4 (2021), 993–1012. https://doi.org/10.1007/s13347-021-00447-6

[36] Joseph Kupfer. 1987. Privacy, autonomy, and self-concept. *American Philosophical Quarterly* 24, 1 (1987), 81–89. https://www.jstor.org/stable/20014176

[37] John Locke. 1690. *An essay concerning human understanding.* Printed for Thomas Basset.

[38] Jie Lu, Anjin Liu, Fan Dong, Feng Gu, João Gama, and Guangquan Zhang. 2019. Learning under Concept Drift: A Review. *IEEE Transactions on Knowledge and Data Engineering* 31, 12 (2019), 2346–2363. https://doi.org/10.1109/TKDE.2018.2876857

[39] Weizhi Ma, Min Zhang, Chenyang Wang, Cheng Luo, Yiqun Liu, and Shaoping Ma. 2018. Your Tweets reveal what you like: Introducing cross-media content information into multi-domain recommendation. In *International Joint Conferences on Artificial Intelligence (IJCAI)*. 3484–3490. https://www.ijcai.org/proceedings/2018/0484.pdf

[40] David E. Melnikoff and Nina Strohminger. 2020. The automatic influence of advocacy on lawyers and novices. *Nature Human Behaviour* 4, 12 (2020), 1258–1264. https://doi.org/10.1038/s41562-020-00943-3

[41] Ingo Mierswa, Michael Wurst, Ralf Klinkenberg, Martin Scholz, and Timm Euler. 2006. YALE: Rapid prototyping for complex data mining tasks. In *12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 935–940. https://doi.org/10.1145/1150402.1150531

[42] Deirdre K. Mulligan, Colin Koopman, and Nick Doty. 2016. Privacy is an essentially contested concept: A multi-dimensional analytic for mapping privacy. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 374, 2083 (2016), Article No. 20160118. https://doi.org/10.1098/rsta.2016.0118

[43] Shaun Nichols and Joshua Knobe. 2007. Moral responsibility and determinism: The cognitive science of folk intuitions. *Noûs* 41, 4 (2007), 663–685. https://www.jstor.org/stable/4494554

[44] Kenya Freeman Oduor and Eric N. Wiebe. 2008. The effects of automated decision algorithm modality and transparency on reported trust and task performance. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 52, 4 (2008), 302–306. https://doi.org/10.1177/154193120805200422

[45] Alexander Pak and Patrick Paroubek. 2010. Twitter as a corpus for sentiment analysis and opinion mining. In *Seventh International Conference on Language Resources and Evaluation (LREC)*. 1320–1326. http://www.lrec-conf.org/proceedings/lrec2010/pdf/385_Paper.pdf

[46] Carina Prunkl. 2022. Human autonomy in the age of artificial intelligence. *Nature Machine Intelligence* 4, 2 (2022), 99–101. https://doi.org/10.1038/s42256-022-00449-9

[47] Daniele Quercia, Michal Kosinski, David Stillwell, and Jon Crowcroft. 2011. Our Twitter profiles, our selves: Predicting personality with Twitter. In *IEEE Third International Conference on Social Computing*. 180–185. https://doi.org/10.1109/

PASSAT/SocialCom.2011.26
[48] Dimitrios Rafailidis, Pavlos Kefalas, and Yannis Manolopoulos. 2017. Preference dynamics with multimodal user-item interactions in social media recommendation. *Expert Systems with Applications* 74 (2017), 11–18. https://doi.org/10.1016/j.eswa.2017.01.005
[49] Antoinette Rouvroy and Yves Poullet. 2009. The right to informational self-determination and the value of self-development: Reassessing the importance of privacy for democracy. In *Reinventing Data Protection?*, Serge Gutwirth, Yves Poullet, Paul De Hert, Cécile de Terwangne, and Sjaak Nouwt (Eds.). Springer, 45–76. https://doi.org/10.1007/978-1-4020-9498-9_2
[50] Piotr Sapiezynski, Avijit Ghosh, Levi Kaplan, Alan Mislove, and Aaron Rieke. 2019. Algorithms that" Don't See Color": Comparing Biases in Lookalike and Special Ad Audiences. *arXiv preprint arXiv:1912.07579* (2019).
[51] Marya Schechtman. 1996. *The constitution of selves.* Cornell University Press.
[52] Marya Schechtman. 2011. The narrative self. In *The Oxford Handbook of the Self*, Shaun Gallagher (Ed.). Oxford University Press.
[53] Marya Schechtman. 2014. *Staying alive: Personal identity, practical concerns, and the unity of a life.* Oxford University Press.
[54] Andrew D. Selbst, danah boyd, Sorelle A. Friedler, Suresh Venkatasubramanian, and Janet Vertesi. 2019. Fairness and abstraction in sociotechnical systems. In *Second Annual Conference on Fairness, Accountability, and Transparency*. 59–68. https://doi.org/10.1145/3287560.3287598
[55] Dusan Sovilj, Scott Sanner, Harold Soh, and Hanze Li. 2018. Collaborative filtering with behavioral models. In *Conference on User Modeling, Adaptation and Personalization (UMAP)*. 91–99. https://doi.org/10.1145/3209219.3209235
[56] Luke Stark. 2018. Algorithmic psychometrics and the scalable subject. *Social Studies of Science* 48, 2 (2018), 204–231. https://doi.org/10.1177/0306312718772094
[57] Luke Stark. 2019. Facial recognition is the plutonium of AI. *XRDS: Crossroads, The ACM Magazine for Students* 25, 3 (2019), 50–55. https://doi.org/10.1145/3313129
[58] Luke Stark and Jevan Hutson. 2022. Physiognomic artificial intelligence. *Fordham Intellectual Property, Media & Entertainment Law Journal* 32, 4 (2022), 922–978. https://ir.lawnet.fordham.edu/iplj/vol32/iss4/2/
[59] Daniel Susser, Beate Roessler, and Helen Nissenbaum. 2019. Technology, autonomy, and manipulation. *Internet Policy Review* 8, 2 (2019), 1–22. https://doi.org/10.14763/2019.2.1410
[60] Charles Taylor. 1976. Responsibility for self. In *The Identities of Persons*, Amélie Rorty (Ed.). University of California Press.
[61] Charles Taylor. 1989. *Sources of the Self: The Making of the Modern Identity.* Harvard University Press.
[62] José Van Dijck. 2013. 'You have one identity': Performing the self on Facebook and LinkedIn. *Media, Culture & Society* 35, 2 (2013), 199–215. https://doi.org/10.1177/0163443712468605
[63] Jan Van Gemert, Cor Veenman, Arnold Smeulders, and Jan-Mark Geusebroek. 2010. Visual word ambiguity. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 32, 7 (2010), 1271–1283. https://doi.org/10.1109/TPAMI.2009.132
[64] J. David Velleman. 2006. *Self to self: Selected essays.* Cambridge University Press.
[65] Giridhari Venkatadri, Piotr Sapiezynski, Elissa M. Redmiles, Alan Mislove, Oana Goga, Michelle Mazurek, and Krishna P. Gummadi. 2019. Auditing Offline Data Brokers via Facebook's Advertising Platform. In *The World Wide Web Conference*. 1920–1930. https://doi.org/10.1145/3308558.3313666
[66] Hongzhi Yin, Bin Cui, Ling Chen, Zhiting Hu, and Xiaofang Zhou. 2015. Dynamic user modeling in social media systems. *ACM Transactions on Information Systems* 33, 3 (2015), 1–44. https://doi.org/10.1145/2699670
[67] Dan Zahavi. 2007. Self and other: The limits of narrative understanding. *Royal Institute of Philosophy Supplements* 60 (2007), 179–202. https://doi.org/10.1017/S1358246107000094
[68] Fattane Zarrinkalam, Hossein Fani, and Ebrahim Bagheri. 2019. Extracting, mining and predicting users' interests from social networks. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*. 1407–1408. https://doi.org/10.1145/3331184.3331383
[69] Qian Zhao, Martijn Willemsen, Gediminas Adomavicius, Maxwell Harper, and Joseph Konstan. 2018. Interpreting user inaction in recommender systems. In *12th ACM Conference on Recommender Systems (RecSys)*. 40–48. https://doi.org/10.1145/3240323.3240366
[70] Sicheng Zhao, Amir Gholaminejad, Guiguang Ding, Yue Gao, Jungong Han, and Kurt Keutzer. 2019. Personalized emotion recognition by personality-aware high-order learning of physiological signals. *ACM Transactions on Multimedia Computing, Communications, and Applications* 15, 1s (2019), Article No. 14. https://doi.org/10.1145/3233184
[71] Indrė Žliobaitė. 2010. Learning under concept drift: An overview. *arXiv preprint arXiv:1010.4784* (2010). https://doi.org/10.48550/arXiv.1010.4784

A APPENDIX STUDY MATERIALS

A.1 Hypothetical Vignette Scenario

Please read the following scenario carefully:

Imagine that you are an active member of a global social media platform. Think of a social media platform that is similar to a handful of prominent examples such as Facebook, Twitter, or Instagram. Imagine that, on this platform, you are an active member and regularly post content. For example, you frequently upload images to the platform. When your friends publish similar posts, you commonly "react" to their posts. Generally, you often consume the content that the platform presents to you in its so-called "news feed".

More specifically, the data you share with the platform includes your real name, your age, and gender. You also share your current location with the platform, your social contacts, your relationship status, the type of device you use, and your activity data: what content you view and click on when you use the platform and at what time you do so. You are aware that the social media platform has developed algorithms that attempt to draw a variety of conclusions about you based on the types of data you share. The social media platform states that it uses such "conclusions about you" in order to show you more suitable content and product advertisement.

Some conclusions may be based on the data you share actively and consciously. For example, when you provide your real birthday, the platform uses this information to show you content that it takes to be suitable for your age group. Some conclusions about you are based on data that you share implicitly, so you may not be aware that you have shared such data about you.

For example, since you share your location data, the platform tries to conclude where you work and live. As another example, the platform also tries to conclude what hobbies you have based on your friends' activities. The platform attempts to conclude your interests (e.g., movies, music, or books you might like) and your behaviors (e.g., what you buy, who you meet). It tries to conclude your religious beliefs (e.g., whether you are part of a religion or an atheist) and your political stance (e.g., whether you consider yourself liberal or a conservative). The social media platform also tries to draw conclusions about who you are as a person more generally; for example, how you might react to certain content, how introverted or extroverted you are, or how sociable you are.

Now, please imagine that the social media platform combines and stores all conclusions about you in your user data profile. Again, the social media platform claims that it needs the content of your user data profile to know what content and advertisement you find suitable. The social media platform generates all of its revenues by showing you advertisements.

To recap, there are two different user profiles on social media: One profile that you use to share posts or share messages, your profile on the social media platform. The other one is generated by the social media platform about you, which will be referred to as your "user data profile" for the rest of the survey. All survey questions relate to your user data profile, not your social media profile.

You were shown a description of a social media platform. You will now be asked questions regarding your personal perception of social media platforms like the one described previously. All questions relate to a social media platform that was introduced to you in the opening text.

Please answer these questions from your own point of view.

A.2 Manipulation Checks (in-text)

(1) Asked prior to vignette text: It is important that you pay attention to this study. Please read the scenario described below carefully.

- *Please confirm this by selecting "Strongly disagree."*

(2) Asked at the end of the vignette text: Please indicate which of the following is true.

My user data profile is:

- *My social media profile that I use to socialize when I log on to the social media platform.*
- *My profile that the social media platform's algorithms generate about me based on the data I share explicitly and implicitly.*
- *I don't know.*

A.3 Survey Questions

7-point-scale, 1 = "Strongly disagree" to 7 = "Strongly Agree," and "I don't want to answer."

Questions were divided into 5 categories and shown to respondents in random order within these categories. Participants did not see headlines of question categories.

Accurate judgments (general)

- *The social media platform is able to draw correct conclusions about me.*
- *I believe that the social media platform is able to know what is valuable to me.*
- *I believe that my user data profile is unique. Other users have a different user data profile.*
- *If close friends and family saw my user data profile, they would be able to identify that it's me.*

Accurate judgments (specifics)

- *I believe that social media algorithms are able to accurately conclude what my interests are (e.g., movies, music, or books I like).*
- *I believe that social media algorithms are able to accurately conclude what I have bought in the past.*
- *I believe that social media algorithms are able to accurately conclude what I will buy in the future.*
- *I believe that social media algorithms are able to accurately conclude where I go.*
- *I believe that social media algorithms are able to accurately conclude who I meet.*
- *I believe that social media algorithms are able to accurately conclude my political stance.*
- *I believe that social media algorithms are able to accurately conclude my religious beliefs.*
- *I believe that the social media platform is able to know where I stand on important issues such as my acceptance of the Covid-19 vaccination.*
- *I believe that the social media platform is able to know where I stand on important issues such as climate change.*
- *I believe that the social media platform is able to know where I stand on important issues such as immigration.*
- *The social media platform is able to distinguish between who I am in private and who I am in social contexts on the social media platform.*
- *The social media platform is able to distinguish between who I am in private and who I am in social contexts when I am not online.*

Accurate judgments (temporal)

- *The social media platform is able to keep my user data profile up to date with my interests, behaviors, and beliefs as they change over time.*
- *My user data profile from a month ago includes conclusions about me that are still accurate today.*
- *My user data profile from a year ago includes conclusions about me that are still accurate today.*
- *After having been an active user on the social media platform for several years, the platform can conclude whether I have changed as a person since I started using the platform.*
- *My entire user data profile tells an accurate story of the life that I have lived since I started using the platform.*

The normativity of an accurate UDP

- *The social media platform should take precautions to make sure that my user data profile is accurate.*
- *The social media platform should double-check my user data profile for accuracy, even if it takes them time or possibly other resources (e.g., money or additional employees) to do so.*
- *The social media platform should collect as much of my data as possible to ensure my user data profile is as correct as possible.*
- *The social media platform should collect my data to generate my user data profile.*

Transparency of data collection & use

- *The collection of my data should be disclosed to me clearly and transparently.*
- *The social media platform should disclose the way they collect and use my data.*
- *I want to know what data the social media platform has used to show advertisements to me.*

Transparency of SMP conclusions about me

- *I want to know what the social media platform has concluded about me.*
- *I am interested in understanding all the conclusions the social media platform has made about me.*
- *It is valuable to me to understand all the conclusions the social media platform has made about me.*
- *I do not care about the conclusions that the social media platform makes about me.*

Transparency of UDP

- *It is important to me that I am aware and knowledgeable about how my personal data will be used for my user data profile.*
- *The social media platform should allow me to access my user data profile.*

Changing my UDP

- *The social media platform should allow me to correct errors in my user data profile.*
- *I am motivated to change conclusions that I think are wrong in my user data profile.*
- *I am upset if the social media platform concludes something about me that I think is wrong.*
- *If my user data profile was made transparent to me, then correcting and maintaining my user data profile would be too tedious for me.*

If I could view my UDP

- *If I had the ability to view my user data profile, I would compare elements of the user data profile to the person that I think I am.*
- *If I had the ability to view my interests (i.e., movies, music, or books that I like) in my user data profile, I would compare them to my own real interests.*
- *If I had the ability to view what the social media platform claims I have bought in the past, I would compare it to what I have actually bought.*
- *If I had the ability to view where the social media platform claims I went in the past week, I would compare it to where I really went last week.*
- *If I had the ability to view who the social media platform claims I have met in the past week, I would compare it to who I met in the past week.*
- *If I had the ability to view my political stance in my user data profile, I would compare it to my own real political stance.*
- *If I had the ability to view my religious beliefs in my user data profile, I would compare them to my own real religious beliefs.*

If I was shown the content of my UDP

- *If I was shown the content of my user data profile, it would cause me to reevaluate who I am.*
- *If I was shown the content of my user data profile, it would cause me to reevaluate my interests (i.e., movies, music, or books that I like).*
- *If I was shown the content of my user data profile, it would cause me to reevaluate what I will buy in the future.*
- *If I was shown the content of my user data profile, it would cause me to reevaluate my political stance.*
- *If I was shown the content of my user data profile, it would cause me to reevaluate my religious beliefs.*

Influence of self-concept (unaware elements)

- *If I could view the content of my user data profile, then the conclusions the social media platform has made about me would have meaning to me.*
- *If my user data profile contains conclusions about who I am that I did not know about, then these conclusions don't have meaning to me.*
- *If my user data profile contains conclusions about my political stance that I did not know about, then these conclusions don't have meaning to me.*
- *If my user data profile contains conclusions about my religious beliefs that I did not know about, then these conclusions don't have meaning to me.*
- *If I was looking for a new interest, I would look into my user data profile for a suggestion.*
- *If my user data profile contains conclusions about my interests (e.g., movies, music, or books that I like) that I do not know about, then these conclusions will likely become new interests of mine.*
- *If my user data profile contains conclusions about what I will likely buy in the future, that I didn't know about, then these conclusions will likely influence what I buy in the future.*

A.4 Demographics

54.3% of participants were female, 43.8% male, and 1.9% defined themselves as other. 69% of participants were between 18 and 35 years old. 56.8% of participants had some form of university education, 33.4% had at least a high school diploma. 50.8% of participants were employees, 18.2% were students. Finally, 92.9% of participants listed their current country of residence as the United States.

B APPENDIX FIGURE 7

Figure 7 shows respondents' beliefs on the influence of the UDP on their self-concept. In particular, we wanted to understand whether participants would attribute meaning to identity declarations in their user data profile (UDP) that they were not aware of. Figure 7 is shown on the following page.

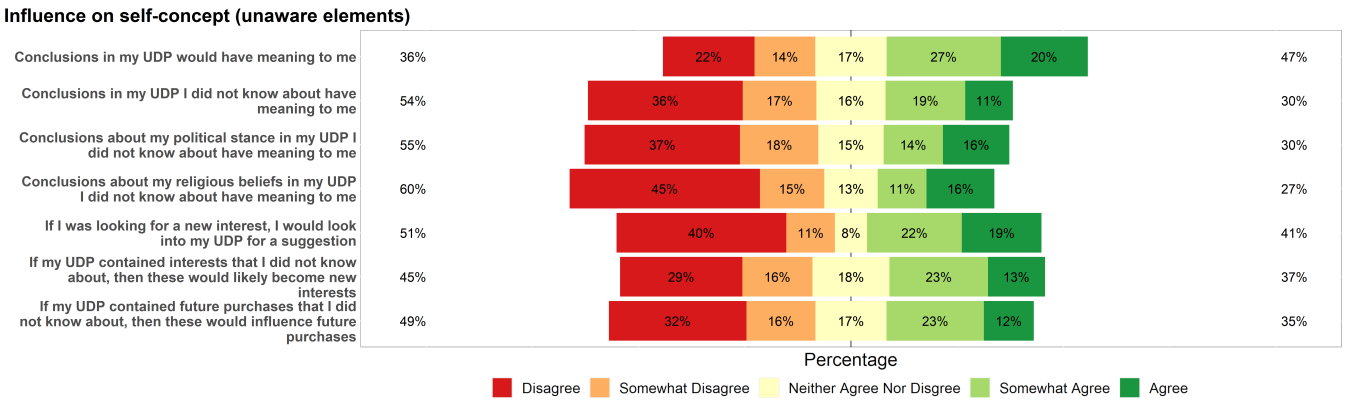

Figure 7: Respondents' ratings were largely divided over the question whether UDP conclusions would be meaningful to them and whether unknown identity declarations would carry meaning for them.