"Becoming Part of Something Bigger": Direct to Consumer Genetic Testing, Privacy, and Personal Disclosure

JENNIFER KING, Center for Internet and Society, Stanford University, U.S.A.

Direct to consumer genetic testing (DTCGT) services where users can identify their inherited diseases and traits, find genetic relatives, and learn more about their ethnic heritage continue to grow in popularity. At the same time, one's DNA is one of the most identifiable, immutable forms of personal information, and sharing it carries risks to one's privacy. What motivates individuals to engage in DTCGT, and what are the implications for information privacy at both the individual and societal levels? This study uses qualitative interviews with ten customers of the DTCGT service 23andMe to explore why they engaged in DTCGT, the benefits they received, their expectations of privacy, and perceptions of risk. It also introduces the use of social exchange theory as a theoretical framework for examining the social dimensions of information privacy and personal disclosure. The findings demonstrate that the participants' assumptions of anonymity, as well as their belief that their contributions to online genetic databases aid the public good, were key motivating factors. The participants were generally unaware of the potential risks to their individual genetic privacy as well as the impact of large scale genetic testing databases on networked, collective privacy. The findings demonstrate that framework of social exchange theory aids in understanding how the form of the relationship affected the participants' decisions to disclose their personal information to 23andMe as well as their perceptions of risk in the DTCGT context.

CCS Concepts: • **Human-centered computing** → **Empirical studies in HCI**; **Empirical studies in collaborative and social computing**; • **Security and privacy** → **Social aspects of security and privacy**;

Additional Key Words and Phrases: information privacy; personal disclosure; direct to consumer genetic testing; DNA; social exchange theory; direct negotiated exchange; indirect generalized exchange; trust; power; networked privacy; institutional privacy; interpersonal privacy; collective privacy

ACM Reference Format:

Jennifer King. 2019. "Becoming Part of Something Bigger": Direct to Consumer Genetic Testing, Privacy, and Personal Disclosure. *Proc. ACM Hum.-Comput. Interact.* 3, CSCW, Article 158 (November 2019), 33 pages. <https://doi.org/10.1145/3359260>

1 INTRODUCTION

Direct to consumer genetic testing (DTCGT), with an estimated 12 million customers in 2018, continues to grow in popularity.[48] For a modest fee, anyone can submit a physical DNA sample (usually saliva-based) to learn more about their risk of inherited diseases, presence of inherited traits, and their probable ethnic composition. Participants may also identify other genetic relatives who have elected to share their DNA with the same service.

Through the use of semi-structured qualitative interviews this study explores the privacy expectations of ten direct to consumer genetic testing (DTCGT) users, as well as their motivations for

Author's address: Jennifer King, Center for Internet and Society, Stanford University, 559 Nathan Abbott Way, Stanford, CA, 94305, U.S.A., jenking@law.stanford.edu.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

© 2019 Copyright held by the owner/author(s). Publication rights licensed to ACM.

2573-0142/2019/11-ART158 \$15.00

<https://doi.org/10.1145/3359260>

genetic testing, perceived benefits of the service, and their awareness of the privacy risks of DTCGT. It also introduces the use of social exchange theory as a theoretical framework for examining the relational dimensions of personal disclosure to an institution (rather than other individuals). In addition to providing insight into motivations for engaging with DTCGT, this research contributes to the literature in social computing by:

- Examining information privacy and personal disclosure to institutions not only as an action with individual benefits and risks, but as one that has consequences for the individual's genetic relatives as well as the broader society;
- Using the framework of social exchange theory, it examines the motivations and benefits from exchanging one's genetic data both as a *direct negotiated exchange* between individuals and a DTCGT company where they receive *direct* benefits as well as an *indirect generalized exchange* where individuals concurrently contribute to an information pool of genotyped DNA from which they benefit *indirectly* (e.g., as a common or public good);

This paper argues that the exchange of personal information—in this instance, a sample of one's DNA—for access to a information-based service is a specific form of social exchange: a *direct negotiated exchange*. In addition to understanding why people electively use DTCGT services and their expectations of privacy and perceptions of risk, a broader research question motivating this work is: how do specific forms of exchange affect individuals' decisions to disclose their personal information to a company or organization?

DTCGT represents a new frontier in the collection of personal information. Unlike most consumer services, which may collect a broader variety of data about individuals, DTCGT collects a biometric that is not only uniquely identifiable and immutable and that can reveal one's inherited traits and diseases, but one that also allows for the identification of others genetically related to the elective donor. The current and future risks to individual privacy are profound, albeit difficult for non-experts to grasp because their impact may be greater for collective rather than individual privacy. But it is precisely because there is a significant disconnect between the public's understanding of DTCGT's risks that the area requires attention from experts to raise concerns now, before consumer genetic testing becomes more widespread. And given that DTCGT operates outside of existing regulatory frameworks governing medical uses of genetic data, identifying vulnerabilities is crucial to guide policymaking before the industry grows too pervasive.

1.1 Background

Direct to consumer genetic testing services such as 23andMe, Ancestry.com, and FamilyTreeDNA, among others, allow their customers to share their genotyped results on their respective platforms in order to find genetic relatives who have also submitted DNA samples to the same service. A select few also provide identification of genetic mutations for inherited diseases or traits. For individuals who have questions about their ethnic or genetic heritage, or concerns about inherited diseases, DTCGT can provide a breadth of answers that few other services, such as doctor-ordered genetic tests, can—and with exceptionally little effort. As 23andMe states on their website: "It's just saliva. No blood. No needles." [3]

Most people seek out consumer genetic testing to trace their ancestry, both to identify relatives as well as to discover or confirm one's ethnic heritage. [50] 23andMe was the first service to offer both ancestry testing and information about inherited traits and diseases. In addition to providing direct individual benefits, some services also aim for a higher calling: asking users to contribute their DNA to genomic research projects. Not only do customers gain highly personalized genetic information about one's self, another motivating factor are these *indirect* benefits: the opportunity to contribute to research studies, further public health research, or the possibility of benefitting

from future medical advances based on genetic data. At least some users are motivated to contribute to commercial genetic testing services because they perceive that by doing so they benefit society rather than only themselves.

For example, the 1000 Genomes project, which concluded recruitment in 2015, sought participation by the public to "provide a comprehensive resource on human genetic variation." The project created a public database of genotype data available to researchers.[29] National Geographic's Genographic Project pitches one of its aims as "to gather and analyze research data in collaboration with indigenous and traditional peoples," allowing individuals to learn about their own specific ancestry, and additionally to contribute to the project's analysis of human ancestry. [40] 23andMe—which is the focus of this study—places research at the core of its product, providing its customers the potential to share their genetic data for research purposes. The company argues that it gives "everyday people the opportunity to make a difference by participating in a new kind of research—online, from anywhere." [2]

While the motivation to contribute to the public good is, of course, admirable, what is less clear is whether contributions to DTCGT services will actually result in a public benefit. Commercial DTCGT services such as 23andMe are privately held with no obligation to use their data for the public good. Although 23andMe has partnered with multiple academic and government research projects, they also partner with private corporations, such as pharmaceutical companies. For example, in July 2018, 23andMe made headlines when they announced that pharmaceutical giant GlaskoSmithKline purchased a \$300 million dollar stake in the company.[18] 23andMe's terms of service state that "you should not expect any financial benefit from 23andMe as a result of having your Genetic Information processed; made available to you; or, as provided in our Privacy Statement and Terms of Service, shared with or included in Aggregated Genetic and Self-Reported Information shared with research partners, including commercial partners." Customers are advised that they "will not receive compensation for any research or commercial products that include or result from your Genetic Information or Self-Reported Information." [4]

While DTCGT offers its customers unique direct benefits as well as the opportunity to contribute to scientific research, the commercial collection and sequencing of genetic data carries with it inherent risks to both individual and societal privacy. While most DTCGT services today do not offer full genomic sequencing, genotyped DNA (which sequences a small sample of one's genetic code) can still be uniquely matched to an individual. Unlike an assigned identifier, such as a U.S. Social Security number, one's DNA cannot (at least, as of yet) be changed, ever. Further, while submitting a genetic sample to a DTCGT company is typically an individual decision, it carries with it consequences both for one's genetic relatives, and increasingly, for humanity writ large. As Erlich *et al* demonstrate in an article published in *Science* in November 2018, with a sample of 1.28 million individuals, at the time of publication "about 60 percent of the searches for individuals of European descent will result in a third-cousin or closer match, which theoretically allows their identification using demographic identifiers." Further, in the "near future" nearly any individual in the U.S. of European descent will be identifiable from these DTCGT databases.[20] In the hands of an authoritarian regime, databases of genetic material can be used for large-scale social control. Already in China, authorities are using DNA-based identification, along with facial recognition technology, to track the country's Muslim Uighur population.[61]

In the U.S., these large consumer genetic databases have already drawn the interest of law enforcement authorities. In 2018, California police created an account using a false identity and submitted a genetic sample from a crime scene to GED Match, a DNA database that allows individuals to submit genotyped DNA data for genealogical research purposes. The sample, belonging to an unknown suspected serial killer and rapist nicknamed the "Golden State Killer," matched a third cousin of the suspect in GED Match's database. Through additional research, a genealogist identified an

individual, Joseph DeAngelo, whom police arrested and charged.[22] The publicity surrounding the case introduced a frenzy of similar investigations, leading GED Match to update their terms of service to allow law enforcement entities to openly submit DNA samples to either identify perpetrators of violent crimes or deceased individuals.[24] In 2019, FamilyTreeDNA faced intense criticism after authorizing law enforcement officials to submit samples of suspected criminals without first alerting their existing customers or changing their terms of service or privacy policy to reflect this change.[26]

These trends are worrisome given that there is only one federal law in the U.S., the Genetic Information Non-Discrimination Act (GINA), that provides protections for genetic data. GINA prevents employers and health insurance companies from using DNA to discriminate against individuals, but there are no legal protections that prevent private companies outside the health information context or other types of insurers from using genetic data for any commercial purpose they wish. GINA also does not provide any protections for other forms of insurance services, such as life or long-term disability insurance. Because consumers can submit DNA samples to DTCGT services directly, rather than at the request of a physician, these companies are not subject to the U.S. Health Insurance Portability and Accountability Act (HIPAA) and the privacy restrictions HIPAA places on patient data.[66] In short, consumer DNA collected by DTCGT companies is presently not subject to regulations in the U.S. that would prevent the companies from electing to share or sell this data with anyone other than an individual's employer or health insurance company. Their only limitation is what they presently specify in their privacy policies, which they are free to change at any time to introduce new uses of their customers' data. Thus, consumers who contribute their DNA to these services are at the mercy of the company to choose to act ethically. This is of course assuming that the company isn't hacked or otherwise loses control of its customers' data.

2 RELATED LITERATURE

As of 2019, neither CSCW nor CHI have published research on the topic of DTCGT. Within the broader scope of ACM proceedings, a 2014 article explored whether consumers were able to interpret the results of genetic-based disease testing [37], and a 2015 article examined the privacy policies of six DTCGT companies.[28] Utilizing the framework of contextual integrity (described below) for their analysis, the authors raised several concerns regarding whether the companies adequately informed customers of potential risks to their genetic data, especially given the unique properties of genotyped DNA as compared to other types of personal information routinely collected online. In short, DTCGT remains an emergent area for research and analysis.

This paper builds upon others' work examining privacy in networked spaces, as well as studies that similarly draw upon social exchange theory to situate the process of exchanging goods and services (and in this case, information) in computer mediated environments.

2.1 Privacy: Interpersonal and Institutional

The study of privacy within CSCW has been largely focused on personal disclosure between individuals in computer-mediated environments such as social media sites. Leysia Palen and Paul Dourish's landmark 2003 article *Unpacking Privacy for A Networked World*, which draws upon the theoretical work of Irwin Altman as a framework for examining privacy in networked spaces, emphasizes "the dynamic and multidimensional nature of privacy," elaborating concepts of defining the public versus the private, identity and the boundaries around the self, and how information technologies disrupt notions of the temporal. [?][9] Multiple researchers have also drawn upon Sandra Petronio's theory of communication privacy management, an extension of Altman's theory, to bridge the gap between physical world disclosures of personal information

to those in digital spaces.[43] Taken together, Petronio and Altman's theories provide a rich framework for understanding the complexity of personal disclosure between individuals in computer mediated social relationships, importantly not only as an individual decision, but also as one that has repercussions for the recipients of individually disclosed information. Several authors have examined privacy issues from this perspective. [16][56][65][64][55]

Another important dimension to information privacy is the relationship between individuals and the companies, organizations, and platforms—institutions—to whom they disclose their personal information. Altman and Petronio's theories do not account for the dynamics between a human discloser and a non-human recipient in individual to institutional relationships, which function differently than disclosure relationships between individuals. For example, relationships between individuals and companies often feature structural power imbalances, such as information asymmetries, or are governed by legal frameworks that frame the terms of the relationship in a lopsided manner. Further, the process of negotiation between information co-owners that is a key feature of communication privacy management is difficult to apply to individual-institutional relationships, particularly given the limited opportunities individuals have to negotiate with institutions.

Within the institutional disclosure context, privacy is typically operationalized as one's ability to practice *selective disclosure*—the ability of an individual to choose to whom she wishes to disclose, what, when, and why. This definition follows privacy theorist Alan Westin's conception of privacy as control over one's personal information, and over the last fifty years it has been instantiated into the Fair Information Privacy Practices and practiced as notice and consent.[62] [23]However, not all privacy violations result from a lack of control, and nor do they necessarily result in individualized harm. Thus, philosopher Helen Nissenbaum's theory of contextual integrity (CI) offers a fuller accounting of privacy in these types relationships (and interpersonal relationships as well). Nissenbaum defines information privacy as based on contextually dependent informational norms, with violations of privacy resulting from normative violations of information flows.[41] When information flows respect contextual norms, individual (and arguably, societal) information privacy is maintained. In 2012, Louise Barkhuus called broadly on Human-Computer Interaction and UBICOMP researchers to draw upon CI in their examinations of personal information sharing practices "for a more nuanced treatment of the notion of privacy within HCI." [11] Potentially because CI has less of a focus on interpersonal interactions and more on societal level norms and systemic applications, there has been less reliance on CI in the CSCW literature. One reason may be that information flows are typically designed at the system level and imposed on users, rather than created from the bottom up. Accordingly, several of the papers that draw upon CI to examine privacy focus on systemic-level privacy issues rather than interpersonal ones.[59][14][54]

Marwick and boyd interleaved these two perspectives in their definition of privacy as a *property of a network* rather than the individual. As they argue:

Networked privacy invokes the constellation of audience dynamics, social norms, and technical functionality that affect the processes of information disclosure, concealment, obscurity, and interpretation within a networked public. If we understand privacy to be about the management of boundaries, networked privacy is the ongoing negotiation of contexts in a networked ecosystem in which contexts regularly blur and collapse. Networked privacy cannot be achieved simply by providing or denying information; it requires meaningful control over the networked contexts in which the information flows. In other words, achieving privacy requires that people have an understanding of and influence in shaping the context in which information is being interpreted. This can be done by co-constructing the architecture of the systems, or it can be done by embedding meaning and context into the content itself. [36]

The DTCGT environment, and 23andMe's platform specifically, includes elements related to all of these approaches to privacy. When users elect to share their genetic identities with other users, they abut against the issues of boundary management and context collapse. When they share their data with the platform itself, as well as with the research component, they do so in a specific context where information-flow norms will inform their expectations of what is the appropriate use of their data. However, this study focuses primarily on the institutional disclosure context between the individuals and 23andMe, acknowledging the explanatory significance of CI as a theoretical basis for examining privacy. The interpersonal elements, while important, are not the primary focus of this research.

2.2 Personal Disclosure and Social Exchange Theory

In order to examine personal disclosure as an act independent of privacy concerns or expectation, this study uses social exchange theory (SET) as the theoretical framework to structure its inquiry. SET provides an empirically validated means for examining disclosure as a feature of an exchange relationship between individuals and companies. Although contextual integrity is an essential theory for interpreting why non-normative flows of information disrupt context specific privacy norms, at the same time CI is limited in explaining why individuals disclose personal information in digital contexts where the norms are emergent, in flux, or when a company's practices may violate contextual norms. Further, a company's data collection and usage practices aren't always obvious or well-understood by its users, and thus trying to explain personal disclosure on the basis of one's privacy expectations or knowledge may foreclose other motivations for disclosure, especially when most people are likely not thinking foremost about privacy concerns or risks when making disclosure decisions. DTCGT services are still relatively new and growing, with a comparatively small customer base compared to more mature online services. Arguably, the norms around many aspects of DTCGT are still emergent, as are the public's knowledge of privacy risks and expectations with data use in the DTCGT context.

Social Exchange Theory (SET) is in actuality a set of theories with roots in sociology and social psychology focusing "on the benefits people obtain from, and contribute to, social interaction." [38] Key to the theory is the recognition that actors exchange goods and services in multiple ways, relying upon monetary exchange and valuation, but also "gift giving, negotiation, and normative understandings of reciprocity," resulting in both economic and social outcomes. [32] CSCW researchers have drawn upon social exchange theory primarily for theorizing the basis of gift giving and similar forms of reciprocal exchange in online communities, as well as on platforms that rely upon trust and other assurances to facilitate exchanges of services among strangers. [33][32][57]. For example, Lampinen and Cheshire used social exchange theory to structure their qualitative interview study of Airbnb hosts to understand the relationship between how financial incentives, intrinsic motivation, and platform assurances contributed to the facilitation of Airbnb hosting between strangers. [32] Lampinen *et al* also used SET in their qualitative case study examining indirect generalized exchange in a geographically-based online community focused on exchanging favors. [33]

While these exchanges are often focused on transfers of goods or services between individuals on computer-mediated platforms, personal information is also a good, one that is often exchanged for access to a service. A SET-based relational analysis incorporates social structural factors that may impact one's decisions to disclose personal information independent of privacy concerns, principally the distribution of structural power between the actors, and its effect on perceptions of trust, fairness and benefits of the exchange relationship. Exchange occurs between many possible configurations of individuals and groups in multiple forms of exchange relationships; for the purposes of this study, the primary form of the exchange between 23andMe customers and the

company is defined as a *direct negotiated exchange*, where the exchange occurs directly between the two actors and that the exchange itself has been *negotiated*, meaning the product of the exchange is agreed upon based on a joint decision process that is known to each actor. While this does not mean that individuals are in fact negotiating agreements directly with the company, it does imply that the basic process by which they provide their DNA to the company is predictable and results in a known outcome (e.g., the customer creates an account, pays a fee, sends in their genetic sample, and receives information about that sample from the company).

The benefit of describing the disclosure of personal information as a direct negotiated exchange is that SET provides an empirical framework for predicting the role relational factors, such as power, play in personal disclosure. The explicit recognition of power in exchange relationships differentiates SET from the basic model of economic exchange, where actors are generally portrayed as on equal footing. It incorporates both the reality of inequality between social actors as well as the influence of social network structure on exchange relationships. It is this inequality between individuals and companies that arguably, as SET suggests, can affect why individuals disclose their personal information to companies, particularly when the disclosure otherwise appears paradoxical.

Finally, while the primary relationship examined here was the exchange between the individual participants and the company, as will be discussed below, there was also an explicit secondary exchange relationship occurring simultaneously: an *indirect generalized exchange*. Generalized exchange includes "indirect reciprocity between three or more individuals;" the resources each individual contributes may go toward other individuals, or a collective group, without the expectation of a direct benefit in return.[15] Generalized exchange in digital environments is usually observed as actions where individuals contribute to information pools or other common resources without receiving a direct benefit in turn, such as writing a product review online, or rating digitally mediated transactions. In this instance, the indirect generalized exchange consisted of elective contributions of one's DNA to the platform's research program, typically made with an expectation that the contribution would result in a widespread societal benefit. These dual relationships add a layer of complexity to the exchange and arguably have direct implications on the participants' personal disclosure decisions.

3 METHODS

This interview study was part of a deductive exploration into privacy and personal disclosure using social exchange theory as the primary theoretical framework to frame the inquiry. The goal was to examine a specific form of exchange relationship: individuals choosing to disclose personal information to companies in exchange for a product or service, which takes the form of a *direct negotiated exchange*. The full study consisted of twenty participant interviews focused on two primary information sharing contexts within the field of personal health (direct to consumer genetic testing and pregnancy tracking using mobile applications) and one general informational context (online search) for comparison purposes between the groups. There were ten participants in each of the two primary contexts, and all twenty participants were asked about their primary online search provider. This paper focuses only on the ten genetic testing participants. While understanding the participants' motivations for genetic testing provided a necessary grounding and context for also exploring privacy issues, this study is not an ethnography of the full DTCGT experience.

3.1 Context Selection

DTCGT was chosen as the context of inquiry because there is a clear exchange of personal information (one's genetic material) for access to a service (health or ancestry information based on an analysis of the genetic material). Additionally, DTCGT involves both current and future risks to one's genetic privacy, as well as the privacy of one's genetic relatives. Of the DTCGT

Participant Demographics (N=10)

Time Used	Over 5 years (5), Under 5 years (5)
Ethnicity	White (8), Asian (1), Hispanic (1)
Education	Some college (1), 4 year college (3), some graduate school (3), master's degree (2), adv. degree (1)
Age	18-29 (3), 30-39 (3), 50-59 (1), 60-69 (3)
Gender	Female (6), Male (4)

Table 1. Demographic Summary of 23andMe Participants

companies in existence when the study began, only 23andMe offered two types of services —both ancestry matching and genetic medical testing —that other companies did not. A series of pilot interviews indicated that customers of ancestry-only sites were primarily interested in tracing their ancestral roots, while 23andMe users were interested in both their ancestry and the genetic health information that only 23andMe offered directly to consumers. While there are valid and significant interpersonal privacy concerns with the use of ancestral data, the direct risks posed by the company's collection of ancestral information was more abstract to the pilot participants than the consequences of exposing inherited medical conditions. The pilot interviews indicated that the medical conditions that the 23andMe DNA analysis revealed included both current and future privacy concerns for respondents (as well as, potentially, both current living and future relatives). Thus, I chose to focus the inquiry only on 23andMe customers.

3.2 Participant Recruitment

The interviews were conducted over the course of five months in 2016-2017. Participants were recruited using two local community focused online platforms. A convenience sample was used though efforts were made to diversify within the sample with regard to age, ethnicity and gender. None of the participants were personally known to the author. Potential participants were screened with a questionnaire that asked how long they had been using 23andMe, and which specific service features they used (*e.g.*, health and/or ancestry matching). The participants were each paid \$30 upon completion of the interview.

3.3 Interviews and Data Analysis

The goals of the qualitative interviews were twofold: to explore the relationship the participants had with 23andMe, including their motivations for engaging in genetic testing, and the participants' privacy expectations and perceptions of risk with DTCGT. The questions were organized around the following themes:

- General questions about genetic testing and specific questions about their use of 23andMe's service;
- Questions exploring the presence of behaviors and attitudes that may be explained by social exchange theory, namely: exploring features of the relationship between individuals and the company, the presence or relative absence of assurance structures, as well as benefits, a sense of trust, and risks;
- Privacy-focused questions examining the mechanics of personal disclosure, identification of privacy harms and threats, understanding the participants' privacy concerns (*e.g.*, potential concerns about data in relation to insurance companies, potential exposure of friends and family), and exploration of privacy expectations and risks with DTCGT specifically.

The participants' elective use of 23andMe was used as an entry point to ask questions about their experience as well as questions probing individual perceptions of their information exchange relationship with the company; their perceptions of the risks and benefits of this relationship; and their expectations regarding the privacy of the data they shared with the company, among other topics (please see the Appendices for more detail). All of the interviews were recorded, transcribed, and coded using structural coding, which according to Saldaña "both codes and initially categorizes the data corpus to examine comparable segments' commonalities, differences, and relationships" that emerged from the interviews.[52]

The coding scheme was developed by mapping the core concepts of the interview protocol into families of specific codes. Coding was a collaborative process; the first round of coding was performed by a research assistant, using a primary interview coded as an exemplar. Emergent codes were then developed based on themes that arose during the coding process that the protocol-based codes did not capture. After the coding was complete, reports were generated for each code, which included specific quotes identified by participant number, and then the reports were grouped based on themes (e.g., all codes related to SET, all codes related to privacy, etc.). Code summaries were created and used as the building blocks of the data analysis, comparing and contrasting the themes within the respondent pool. Themes were weighted by frequency to highlight which were more widely expressed. Themes with less support, including an occasional outlier, were included in the analysis but were indicated as such, so that their influence would not be overstated. For example, on the theme of the contribution of genetic testing to science and public health (which emerged from questions about sharing one's information on the company's platform, specifically regarding opting-in to the company's research component), all ten participants discussed this topic, with the most discussed sub-themes relating to: the need for public contributions to large health data sets, the benefit of contributing genetic data to medical research generally and genetic research specifically, and the ability to contribute to 23andMe's research being a primary driving factor in selecting the company. One participant noted that she thought the correlational survey questions asked by 23andMe were "bad science," but she was the only participant to make this point during the study. The themes discussed in this article were those with wide support among participants unless otherwise indicated.

3.4 23andMe Company Background

23andMe was founded in 2006 and launched its testing service a year later. The company's growth has been rapid as of late; the number of customers passed the 100K threshold in 2011, the 1MM threshold in 2015, and over 2MM in 2016. 23andMe has had a tumultuous relationship with the Food and Drug Administration, the federal regulatory body that oversees the company in the U.S., regarding how they provide results to customers about inherited diseases. After giving detailed results to consumers regarding their probability of developing specific inherited diseases, the company was required in 2013 to cease providing consumers with inherited disease results, limiting them to only providing ancestral data.[46] In 2015, the company was allowed to provide inherited health results again, albeit in a more restricted form than previously, and for fewer genetic conditions. In April 2017, the FDA gave the company explicit approval to provide users with genetic risk data for ten specific diseases.[58] The 2013 cease-and-desist was widely covered by the media and 23andMe communicated with its existing users about the change. As such, all the participants were asked if they were familiar with the event and how it both impacted them directly as well as whether it had any impact on their relationship with the company.

3.4.1 Genetic Testing Process. In order to obtain DNA sequencing results from 23andMe, customers purchase a DNA collection kit (priced between \$99-\$199, depending on when it was purchased).

The primary exchange in the 23andMe users' relationship is initiated during the sign up phase through the submission of a DNA sample to the company—a vial of one's saliva. After receiving a testing kit in the mail, new customers are asked to sign up for an online account and fill out an online survey disclosing detailed information about one's health history and family background. Though users are not required to disclose any additional information after this exchange, they are incentivized to continue both their relationship with the company and additional disclosure through ongoing survey participation, new research studies, and if the participant has opted into the company's genealogy platform, potentially new family member matches.

Approximately six weeks after they return their sample, customers are contacted with their results via email, which they view on the company's website. Customers are shown information about their inherited traits and diseases, as well as their probable ethnic composition. If one wishes to share their genetic information on the company's platform to identify potential relatives, they can opt-in to do so using the company's DNA Relative Finder service; they can also opt-in to compare results with immediate family members if they know any to have also submitted a sample. 23andMe also partners with research universities, pharmaceutical companies, and non-profit organizations and has received federal grants from the National Institutes of Health to use their customers' genetic data for research. The research is based on a combination of DNA and self-reported data; when 23andMe customers opt-in to the research component, they are asked to complete short, discrete surveys ('Quick Surveys') on an ongoing basis. Their answers, along with their anonymized DNA samples, are made available to the company's research partners for associative analysis.

4 FINDINGS

Several key findings emerged from these interviews:

- Within the framework of social exchange theory, the participants' relationships with 23andMe can be in part explained by the model of a *direct negotiated exchange*. Interviews revealed that these relationships were to a great extent: (a) transactionally focused; (b) provided a clearly understood direct benefit to the participant; and, (c) relied on assurances to mitigate risk and uncertainty.
- In addition to the direct negotiated exchange, the participants were also simultaneously engaged in an *indirect generalized exchange* with 23andMe through their elective contribution of their DNA to 23and Me's research programs, where they assumed that they were contributing to a greater societal benefit through sharing their DNA. Participants' contributions to this indirect generalized exchange deepened their commitment and motivation towards their decision to participate.
- Overall, these voluntary DTCGT users had few immediate privacy concerns about the use of their DNA, though some articulated general or future concerns. The participants expressed few privacy risks with the use of their DNA; most assumed that the company was interested in their data in the aggregate, and not in them as individuals, even though nothing would prevent any DTCGT company from targeting individuals.

4.1 Motivations for and Benefits of Direct to Consumer Genetic Testing

The participants all reported signing up for 23andMe with the express purposes of obtaining genetic information about inherited diseases as well as personal ancestral information. They described their impressions of the service in multiple ways, but all in positive terms: "fun," "cool," "unique," "special," "interesting." Most cited the health and medical offerings as a key reason for participating in genetic testing: wanting to know how their genetics could predict personal health conditions and risks for diseases.

Some participants cited gaps in ancestral memory for motivation: an adoption in the family, divorce and being raised in a single-parent home, or being the only remaining living family member. As one participant noted, "With the speed of research and the advancement in technology, it's always nice to have additional information on your health, because you just never know. I don't mind sharing a little bit of myself out there if I get additional information. It's always nice to be a little preventive, because again, my family aren't the best at recording health history, so I have no idea what runs in our genetics or what doesn't" [P20] A few participants were motivated by the potential of locating unknown living family members: "Now the reason I got into it, this is complex. My mother was adopted, so she never knew her parents, her real bio parents. My bio dad, and my mother divorced either just before, or just after I was born. I never knew anything about him. . . I've always wanted to know something. I always had this sense that maybe I have brothers and sisters, or relatives out there that would help me figure out who the hell I am. I went ahead and sprung whatever it was, \$199, and thought, 'You know this is a good—this'll be helpful.' " [P12]

4.2 Participant Impressions of 23andMe

The participants' impressions of the company were overwhelmingly positive. The most common positive factor mentioned was the participants' belief that they were helping the company to solve problems for the public good. "What's nice is that from what I read, you're part of a bigger scientific study, right? In some ways it feels good to help further science research." [P20] While the main motivation for signing up was to gain personal insights about themselves, most of the respondents noted that their contribution to 23andMe's medical research influenced their decision to participate in genetic testing, as well as their selection of 23andMe over other genetic testing companies. "[Compared to Ancestry.com], 23andMe is more of just trying to understand the human genome more. It's more of big question things rather than let's find your lost grandparent. I think that the [company's] icon points to that, too." [P13]

As another participant mentioned, quite on point, "[i]t feels, I guess, more like a community garden where everybody just goes and they do their own thing, but it's all together. . . we're all cooperating, and we're all getting something out of it." [P1]

Other such generally positive impressions were driven by participants' perceptions that the company was well-funded and well-established, as well as by the visual design and good usability of the website. "It's fun. Yeah. It's bright, bright colors. It's, what, pink and green and fun? It's really not medical. They do a very good job of not seeming Big Brother-esque and making it like, 'Hey, find out more about yourself.'" [P19]

Nearly all of the participants were aware that 23andMe had been under regulatory scrutiny with the U.S. Food and Drug Administration (FDA), though most had only a vague understanding about why. However, the FDA's intervention did not tarnish the company's reputation in most participants' view. In general, the respondents who were members prior to the intervention had a positive view of the company and a negative view of the intervention because it led to their having access to less information than they had previously. "I thought it was stupid. I liked having that access . . . [y]eah. I don't agree with that ruling. I think people should have access to their medical information." [P17]

4.3 Social Exchange Theory and the Company-Customer Relationship

The questions about the nature of the relationship between the participants and the company were intended to draw out insights about the core of the exchange and the participant reasoning behind the disclosure: what did participants believe they are getting from the relationship, and what did they think they are giving up? What did they think the company gained from the relationship?

In discussing what they *gained* from the relationship, the participants spoke of the insights they learned about themselves, their genealogical heritage, the heretofore unknown relatives they were connected to, and their genetic health information, which was of particular value to those concerned about inherited conditions. As a participant mentioned, for example, "[T]hey have told me a little bit about my body and my ancestry and my markers for certain conditions. I feel like that was almost a gift... [t]hey told me something about myself, and I essentially gave them very little... I just hope that it can be used for something good." [P19] Overall, respondents reported receiving what they anticipated from 23andMe; the exchange conformed to the expectations they had when they first entered the relationship with the company, providing support for classifying this exchange as a direct negotiated exchange.

The few unexpected outcomes participants reported were focused on the genetic reporting results not conforming to some of their pre-held expectations, or on disagreements with the company's methods for correlating self-reported survey results with genetic tests. Most were impressed by what they learned about themselves in exchange for the cost and a vial of spit: "I didn't really know what to expect in terms of the value that I'd get back, but what I did get back was cool enough that I—yeah, it was worth 100 bucks and the sample." [P15]

When discussing their *expectations*, respondents referenced their ability to exert control over their data (e.g., opting-in to research studies, rather than opting-out, and sharing data on the ancestry platform were optional), their expectation that their data was anonymized when analyzed or used for research, the lack of any negative consequences or intrusions (thus far), and the belief that their data was being used for public good. "I feel like they've given me a lot of personal agency over opting in to things, that I get to choose what I wanna share, what I wanna tell them." [P19]

Despite there not having been active negotiations between the participants and the companies in a classical sense—no one sits down with a representative from the company to jointly negotiate the terms of disclosure—they are negotiated in the sense that one partner (in these contexts, the company) offers the terms and the other partner (the individual) accepts or declines. While many of us may find this form of negotiation to be unsatisfying, given that the terms are 'take-it-or-leave-it,' rather than an actual give-and-take negotiation, it is reflective of both the power dynamics and existing social norms in the relationships individuals often have with information-based companies today rather than of 23andMe specifically.

4.3.1 *Benefits and Fairness.*

"It's not really my information that gives them the power. It's the information of everybody's." [P13]

The participants were asked who benefited from these exchanges and how fair they believed them to be. While self-knowledge was both the primary motivator and benefit for participants, the key secondary motivator was their contribution to an *indirect generalized exchange*—the belief by most participants that while they were gaining something for themselves, they were also making a contribution to health research that may or may not benefit them directly but had potential value to the public at large. A core assumption of SET is the expectation that "actors behave in ways that increase outcomes they positively value and decrease outcomes they negatively value." [38] Benefits in a direct exchange relationship flow bilaterally (although not always equally), and the participants were asked how fair or beneficial they thought the relationship was to them, as well as how they thought the companies benefited. Another feature of exchange relationships is fairness: how fairly are resources distributed in relation to the relative power held by each of the actors? Did the participants believe they had access to other resources or companies who could provide them with comparable benefits? Did they believe they had the capacity to leave the relationship and find a comparable alternative? Obtaining insight into these dynamics provides additional context to

the nature of the relationship and the reasons why participants choose to disclose their personal information.

Overall, the respondents felt they obtained a considerable direct benefit from their genetic test primarily due to its novelty and uniqueness—the combination of the results that 23andMe provided could not be obtained elsewhere at the time the interviews were conducted. Further, respondents found their personal results both interesting and salient, especially those respondents who joined before the FDA intervention, when more specific inherited disease information was available. Finally, participants mentioned the benefit of sharing information in terms of its personal utility, such as the enjoyment that they got from answering survey questions, gaining information that previously only their doctors may have been able to obtain for them, as well as the ability to provide helpful information about their genetic predispositions to their doctors.

The participants also reported indirect benefits from their relationship with 23andMe. For example, while the participants understood the company benefited from direct revenue received from its customers, their discussions also included a different benefit—the gains to both the company as well as the participants from the participants' generalized contributions to the genetic research pool. There was, in other words, an externalized benefit from this transaction—that of contributing their genetic information in order to advance research and the public good.

One participant summarized his belief on this topic:

"I don't know that *I* offer them anything. They've already got my money, of course, and sometimes they have questions that pertain to some of their medical research, and when they do, I'm fine with answering those. I think it's good on a more global, rather than just me specifically. I think it's really good to have the kind of genetic information databases that they have been accumulating. I think it's good for the medical researchers, among other things. I think it's good for people to know that, "Yeah, you've got X number of people have that same potential for problems, or the same potential for success... I think in any medical study, the larger the sampling, the more accurate the results are gonna be, so I'm just very willing to participate in that. Whether I get any personal benefit out of it or not remains to be seen. I think it's a good thing for people to participate in this kind of thing for the future, really." [P12]

This generalized exchange was also explicitly mentioned as a key benefit for the company (after direct revenues from testing kit sales), but not cynically. Participants believed that the exchange was fair, and that the company could not fully benefit unless it continued to increase its scale and recruit more customers. Most respondents appeared to earnestly believe in the company's mission, and understood that the value and potential of 23andMe's service increases as the company expands its database. As one respondent remarked, "the data I get is probably worth more than my one particular sample of DNA. To them, an aggregate sample of the population is way more valuable than just my one result... they get this large, pool of data that's worth way more than what my results would have even been. Then you can start to draw some interesting things about the population. The real value comes from that." [P16]

Another participant suggested, "I also feel like I'm contributing to a greater good. Being able to do research across a number of individuals. I think that's important, because especially with women, you know? We don't have a lot of medical information about women because so much research was done on men. I figure I'm going to put my genes out there." [P17]

While most participants characterized the exchange as equal, given that participants paid for the service and received a product in return, several did reflect on the fact that giving up their DNA

was likely a non-revocable act.¹ Another participant noted that while she felt that she could engage with the company if needed, she didn't think it was possible to have her DNA removed from the service: "I feel like I could give them feedback. Whether I could make them, let's say, get rid of my data, and erase it from their server, and dispose of whatever copies of my material they have, I don't think they would do that." [P16] It is notable that the majority of these reflections did not lead to a critical assessment of the relationship or the power balance between the service and the participants.

Based on the participants' responses, in some interviews additional questions were posed about the use of their data by 23andMe's research partners, such as pharmaceutical companies, where the participant's data could help to develop future drugs and reap profits for both companies but yield no direct benefit to the participant.² These participants continued to maintain that as long as someone could benefit from the drug that this example of use of their data was acceptable. The commitment to this narrative of contributing to public benefit was strong, and even suggesting scenarios of 23andMe directly profiting from individual data in some interviews didn't alter this commitment. As one participant noted in response to a question asking whether they cared if their data was used to benefit pharmaceutical research, "[t]he reason I filled out that kit—even though there are other cool aspects to it, I wanted to find out more about my genetics. Once I got that payoff of 'Oh, I can log in and see what percent Neanderthal I am,' once I got that, I don't really care what they do with the rest of my data." [P19]

Another said: "I never thought of that. Do I have a concern? Well, let's say that a pharmaceutical company wanted to do some research on what kinds of genetic traits are found in, I don't know. I mean, I think that's a good thing. Right? It benefits us. I mean they're making money, but at the end of the day, we're gonna benefit from that." [P17]

In sum, the combination of a highly unique and personalized product that is both painless to provide and relatively affordable to obtain, along with the option to contribute to scientific research that "sometime in the future you or your family may benefit indirectly from" created both a powerful incentive and both direct and indirect benefits for the participants.³ As one participant aptly summarized: "I feel like I'm not really offering them very much because there's so many other people doing it, but I know if no one did it, then, it wouldn't be good. Yeah, I guess, I'm offering them something good, but it's just very painless for me to contribute it, so it's no sacrifice." [P14]

4.3.2 Trust and Assurance. Within social exchange theory, relationships are understood to be characterized by *risk* and *uncertainty*, and the negotiative aspect of a direct negotiated exchange is meant to mitigate these factors. *Assurances* also mitigate risk and uncertainty in exchange relationships by providing an external or third party mechanism to assure or guarantee the exchange. According to Molm, "[m]echanisms that provide assurance include legal or normative authorities that impose sanctions for violations of agreements or failure to fulfill one's obligations, guarantees such as collateral that protect against loss, warranties that assure certain standards of quality, and so forth." [39] As such, negotiated exchange relationships are marked by less reliance on trust because assurances supplant the need to build trustworthy relationships.

Questions of trust were explored by asking the participants what factors made them comfortable using 23andMe. The visual design of the website and its information presentation were the most cited. Seven of the participants commented on the website's design choices, calling it "bright" and

¹It appears from 23andMe's [Terms of Service Agreement](#) (visited 3/15/19) that if you use the service but do not opt into the research and biobanking aspects, you may be able to request deletion of your account and fully revoke your genetic information from the service.

²Note that this was before the company announced their partnership with pharmaceutical company GlaxoSmithKline.

³<https://www.23andme.com/about/consent/>; visited 3/15/19.

"fun" and nonthreatening. One participant stressed the minimalist design of the logo and the professional presentation of the test kit, expressing that the test kit felt "approachable." The presentation of information on the site also acted as an assurance. Participants cited the well-organized, clear, professional, and clean presentation of the company's website and the information they provide about their intentions for the use of their customers' DNA as improving their trust in the company.

Assurance based on personal relationships was another factor, with four participants having received a personal recommendation from a family member or friend. In one case, a close relative bought kits for everyone in the family: "She was so enthusiastic about it and about what it could show us that—she was really excited, and I think that enthusiasm rubbed off on me a little bit." [P19] Two participants mentioned the business relationship between Google and 23andMe as an assurance, citing their belief that Google was trustworthy as the basis for their similar assessment of 23andMe. Others also mentioned that they thought 23andMe had a good reputation, which provided a foundation for fostering trust.

The ability to exert control over personal disclosure acted as an assurance, albeit in different senses and interpretations of the word. Participants cited the option to share their DNA matches or to not share at all. Others deliberately limited the information they shared with the service as a strategy for limiting personal risk. One participant specifically noted that he believed that the information that 23andMe gathered was limited in comparison to comprehensive medically administered genetic tests, making participating in 23andMe's services less risky. Others expressed their reliance on the company's data anonymization policies, assuming that it would be difficult or impossible for anyone else to trace their genetic information back to them individually.

Formal policies and laws also played a role, perhaps more with 23andMe than most sites given the novelty of the product. Unusually, over half of the participants mentioned reading or skimming the company's privacy policy. One mentioned reading the policy because of a specific issue that was of concern to other participants as well: "My only concern, again, would have been that some of this information might have been shared at some point with an insurer that I might be trying to get insured by." [P18] When asked, half of the participants responded that they believed there was a law that protected their data, though only one knew specifically about GINA (two others incorrectly thought HIPAA provided protection). The other half were uncertain or didn't know: "It was an act of blind faith, really. Unlike me. I was just fascinated by the whole concept. Is there a law?" [P18]

When asked to whom they might complain if they had a negative experience, half of the participants mentioned the company directly. Others said they expected that there was government oversight, but only one mentioned an agency (the FDA) by name. Several participants said they would take to social media platforms to harm the company's reputation, and others mentioned relying on the legal system, such as through a class action lawsuit. "Well, if it were really, truly, large, devastating stuff that warranted a class action lawsuit type of thing, where it's like, "They need to change privacy laws because this is so bad," then, yeah, I'd want to get in on that little dance of 'legal battle.' "[P16]

4.4 Privacy: Personal Disclosure, Risk, and Trade-Offs

After focusing on their relationship with 23andMe, the participants were then asked a series of questions about the types of personal information they had disclosed to the company, and then about potential risks and trade-offs.⁴ The goal was to elicit participant perspectives about

⁴In keeping with critiques that argue that asking interviewees explicit questions that may prime privacy concerns, I did not use the word 'privacy' in this line of questioning.[13]

these themes without explicitly driving the conversation into privacy issues. However, several participants introduced the topic by name on their own accord.

4.4.1 Personal Disclosure. It was a given that all of the genetic testing participants shared their DNA with 23andMe, as well as their contact information (name, address, email) in order to purchase and receive their testing kit. Thus, the participants were asked about other forms of disclosure, specifically whether they: completed the intake health profile; opted into research studies (and thus voluntarily filled out additional "quick surveys," which is the primary method for research participation)⁵; and shared their genetic profile on the company's DNA Relative Finder service. In response to these questions, most reported having completed the health profile, and six of the participants said they had opted into 23andMe's research platform, while the other four either did not opt-in or were uncertain. Many also discussed the societal importance of genetics research and the necessity of large sampling to enable its success. "I think everyone should contribute their information to this database. That's the main thing. I understand that people have privacy concerns, but in this case I don't have any security or privacy concerns about 23andMe." [P17]

Four participants did specifically cite their privacy concerns as a reason that they limited their sharing with the company. Only one participant did not choose share his data with the DNA Relative Finder platform. The other nine had, with a wide diversity of experiences and outcomes, including: comparing data with known relatives (like a parent or sibling); finding distant or unknown relatives; and, exploring unresolved questions around their paternity. Almost all of the participants expressed little concern about giving the company their personally identifiable DNA (potentially, for eternity), in part due to the minimal effort required to submit a sample. The test requires filling a vial with one's saliva, which is easy to produce and infinitely renewable. "I felt like what I gave them was next to nothing, although a lot of people I talked to felt like I gave away the most dangerous information I could've given away." [P1]

Overall, because the participants had highly positive and stable relationships with the company, little or nothing occurred prior to our discussions to challenge participants' core motivations or cause them to question their choice to share their data with 23andMe. As discussed previously, the fact that all DNA-based sharing on the platform requires affirmative opt-in drove a high degree of comfort with the process for participants: "I think by opting in to so many things, I felt like they weren't gonna do a little bait-and-switch, like that if I opted in to this thing, all of a sudden they were gonna do something else. I just remember opting in to so many things that it seemed like everything was very specific. I felt like it was relatively risk-free for what I did opt in for because I thought that there wouldn't be a huge—like that they wouldn't do something else with my data." [P19] Further, the information requested, while extensive, did not violate the participants' sense of contextual integrity as it appeared to respect the normative boundaries of the service (e.g., related to one's health).

In probing participants' expectations of whether the relationship was a one-to-one relationship, or if the company was sharing their data with other parties outside of the research context, several of the participants said they believed that some amount of their personal information was kept between them and the company, but what exactly that information was, varied. Some didn't think that their data was being shared with marketers; one noted that if he discovered that his information was shared with commercial entities he would be very uncomfortable and would want to end the relationship, though he worried that there was no way to take back his information: "No, well God, I hope not. If they did [share], then I think I'd drop [them]. Of course, they have my

⁵23andMe's primary research method is to use self-reported survey questions to associate genetic data with participant responses. At the time of writing, the company also had three disease research programs in progress, which require a diagnosis to join. See <https://www.23andme.com/research/> for more information.

information already, so it just wouldn't make a whole lot of sense to drop them. I would be very much uncomfortable with that if they were sharing it [for] commercial [purposes]." [P12] Others who didn't believe that their personal information was being shared suspected the company was making money off of selling aggregated genetic information, or that insurance companies were funding some of the company's research. One didn't believe her information was being shared at all, saying that she trusted the company: "I had a level of trust in their integrity... I don't think I had any misgivings at all." [P18]

Interestingly, most of the participants said that they didn't consider their DNA to be particularly sensitive personal information. Several of the participants also emphasized that what they exchanged with 23andMe was, from their perspective, just data, and not deeply personal. As one participant phrased it, "It's not personal what I'm giving them at all. It's just like little pieces of numbers and letters." [P14] One participant joked that the data he provided wasn't enough for the company to make a physical clone of him, and thus the risks to him were minimal.

Of course, given that all of these participants were voluntary users of 23andMe, this is not a surprising finding in itself, though participants' rationales for why their DNA wasn't sensitive varied. Some suggested the company was getting useful information from aggregated information and that the individual profiles weren't exhaustive or even necessarily accurate. One noted that your DNA doesn't catalogue your past transgressions—highlighting the idea that while DNA may express something about you, it doesn't capture who you *really* are: an amalgam of your thoughts, emotions, and actions. Another participant echoed the idea that genetic information was a descriptive but not holistic view of her, adding that she was unconcerned about keeping her genetic information private because her DNA wouldn't grant someone access to the things she did consider to be private:

"That's almost like a science fiction movie a little bit. I've given someone my genetic information. I don't think I ever really thought of it that way, that somebody has a little vial of [my] genetic information sitting somewhere. I don't think I've ever really looked at it that way because that genetic information, while it does show characteristics about me, it doesn't show me. I feel like there are enough safety—safety might not be the right word. Security. There are enough links that it's not like by getting that data, by getting my sample, they can also get my bank account information and my address and all those things. I feel like there is enough of a separation that my genetic information is just one in a million samples. While it describes me, it doesn't describe the things that I try to keep personal to myself. I see it as being one small part of a really large sample. I leave DNA traces on everything I touch. I guess I don't see it as that... personal." [P19]

4.4.2 Emerging Social Norms, Perceptions of Risk, and Trade-Offs. DTCGT is such a new phenomenon that there are arguably no extant notification norms—and minimal to any established social norms in general—regarding sharing one's genetic testing results on an online platform and potentially connecting to heretofore unknown relatives.^[51] However, it would be presumptive to conclude that voluntary DTCGT users are somehow significantly less concerned with privacy than anyone else. There are several influential factors at play here that mediated the participants' privacy concerns that will be discussed in more detail below, including their expectation of anonymity on the service, and their belief that contributing one's data aids scientific research. However, one factor that does merit additional discussion is that of assessing risk. The potential risks posed by DTCGT that concern privacy advocates and researchers are not widely publicized, and the respondents in this study were generally unaware of them.^[63] DTCGT services are not required to tell their

customers about potential risks, unlike medical services and procedures which may at least raise potential risks while obtaining informed consent.

In order to understand the full calculus of disclosure in these relationships, the participants were asked about the risks they felt were attached to these disclosures, and what they believed they were trading off in order to use these services. Overall, the participants identified few immediate salient risks, evinced high trust and confidence in the company, and identified few trade-offs they felt they had to make in order to use the service. As discussed earlier, because all genetic disclosures on the platform require an affirmative opt-in rather than an opt-out, there was little focus by the participants on privacy controls, though some participants reported deliberately disclosing less information than they were asked for.

When asked, almost all of the participants mentioned some potential risks with genetic testing, though none appeared to believe these risks were either immediate or salient to them. The risks the participants articulated were based on violations of context, such as sharing or selling the respondents' personal information to third parties. Governments and hackers also posed risks through the discriminatory or unauthorized use of personal information. The participants also identified social risks posed by unauthorized disclosures, such as embarrassment or exposure of private information. None mentioned any concern with their DNA being used by any entity other than 23andMe to identify them. One participant (P15), who had a PhD in bioinformatics, brought a technically informed analysis to his assessment of risk: "You could one day think, 'Maybe someone will come up with a way to target you.' It's really far-fetched given today's technology, but with the whole CRISPR/Cas9 targeting enzymes, I'm not sure what havoc could be wreaked in the future. There's some bio privacy concerns, but I don't know. Right now, it just seems so far-fetched that I'm not worried at all." [P15]

The most frequently suggested risk was that of insurance companies acquiring one's genetic information and using the data to discriminate on the basis of having a genetic marker for a disease, despite the existence of GINA. Others mentioned government control: one, as a Medicare recipient, was mildly concerned with the government gaining access to his DNA, and another referenced actions taken by Germany in the 1930s, noting that the U.S. government could potentially use the data for future racial discrimination. Futuristic fears around cloning and personally identified targeting were suggested, with one participant specifically concerned with genetic data feeding into a government controlled social credit system like that being developed presently in China.^[61] Some felt their anonymity was at risk, but potentially mitigated by the large scale of the database. One participant believed that already there was no realistic guarantee of anonymity, given that genetic testing threatened closed adoptions and could reveal previously unknown half-siblings. Another participant mentioned the potential for socio-familial conflict that genetic testing can raise, as well as the fear of learning you may have a genetic disease. "No, to me there's no risk. People who have histories that they might not know about. That's probably the people that are at risk the most. Emotional risk or maybe that one test that does tell you that you're more likely to get cancer or something. Probably those are the kinds of risks that you have to face, but it's kind of important. You should probably know those things about yourself anyway, so to me, no." [P13]

When asked to consider what they were giving up—trading off—to use these services, the participants (not surprisingly, since they were all current users of these services) felt that the trade-offs were minimal or were more than justified than the benefits they gained. That calculus seemed unlikely to change absent a profoundly negative experience that creates greater liabilities than benefits. The participants suggested very few trade-offs they felt they were making to use 23andMe. Several participants emphasized that any potential risk of contributing their DNA to the company was more than outweighed by the direct benefits they received in turn, as well as the potential of their contribution to research: "I'm not bothered by it because I don't feel that I'm

being used. I think it's science. They're trying to gather information, and that's okay. I'm happy to contribute to that. I paid \$100.00 or \$90.00 at the time. Doesn't bother me at all. I got some test results that no doctor would ever give me, probably wouldn't have shared with me if they had ordered." [P18]

Another participant noted: "I think it's a pretty good trade. They're getting data and being able to sell the data. I try not to give them any more than absolutely necessary. I know they're collecting everything. I figure that's okay. I haven't seen anything sinister that I would worry about. I think it's kind of a fair trade on a certain level." [P12]

Additionally, one suggested that: "In the one-to-one exchange, I think, I get a fair trade; but, because they're a company that's operating on this big scale—they get this large, pool of data that's worth way more than what my results would have even been. Then you can start to draw some interesting things about the population. The real value comes from that." [P16]

5 DISCUSSION

The growth of DTCGT services presents an example of how the future of information privacy in networks may develop—a future in which individual choices can have broad repercussions for the individuals who disclose their own data as well as the privacy of others within their network (in this case, a familial network). In DTCGT systems not only are the individuals who electively participate affected, but also everyone else who is genetically related to those individuals within several degrees of connection. Sharing one's genetic information means, in other words, that information about others will be also shared through uniquely held DNA whether or not those genetic relatives have consented, let alone whether or not they are even aware of their relative's decision to participate (or are even aware their relative exists).

One person's choice also belies implications for future others, since anyone who contributes today to a DTCGT database who might have children (or have siblings or cousins who may have children) in the future has included a portion of their future kin's genetic information in these databases. The privacy implications of these services thus move beyond extant consumer privacy threats such as inferences made through behavioral data. These network-based risks are amplified by the use of uniquely identifiable and immutable biometric data, raising concerns similar to the use of such biometric-based technologies as facial recognition or voice printing. It opens up the possibility of a future where private companies, if unrestrained by law, may use your genetic data for any purpose they wish.

DTCGT raises a multitude of interpersonal privacy questions as well, illuminating genetic connections between individuals that many may have desired to never be revealed: birth parents who gave up children for adoption; sperm and egg donors who assumed anonymity; individuals who discover their parentage is not what they thought; who discover unknown siblings; or even unknown offspring.[45] Allowing a private company to act as a custodian for one's identifiable genetic material indefinitely poses multiple risks in today's unregulated consumer data marketplace, including being subject to requests from law enforcement, theft by hackers, and the potential for undesired commercialization.[53] [30] While the identifiability of relatives through DNA may pose minimal risk in typical family relationships, stories continually emerge about the discoveries people made through DTCGT about relatives they knew nothing about. For example, one woman unexpectedly discovered she was fathered by a sperm donor, and had at least three half-siblings, after she shared her genetic profile on 23andMe's platform.[49] Others have discovered that they shared the same donor father: a fertility doctor who lied to his patients and impregnated them with his own sperm.[67] One man, conceived through sperm donation, recently chronicled his effort to meet and photograph the 32 half-siblings he discovered he had who all share the same donor father. While in his case he located his half-siblings through their shared sperm donor ID number, other

donor-conceived children have discovered half-siblings through DTCGT sites.[10][17] Some of the participants in this study shared similar stories, not all with positive outcomes—occasionally these discoveries led to turbulence and strife in families as secrets emerged that the secret holders never intended to be known. While most of these revelations no longer carry the same social stigmas as they may have in the past, they may still be entirely life-changing, for better or worse, for those who experience them.

Yet, despite the gravity of these issues, and their risks to both individual and collective privacy, the participants in this study were generally unaware of them. As mentioned earlier in Section 3, these same participants were also interviewed about their use of online search engines and their privacy expectations of their aggregated search queries. Overwhelmingly, this group was far more concerned about the privacy issues implicated by the collection and use of their search data than of their genetic data. As one respondent described it (in reference to her use of Google, her primary search engine): “Oh my gosh, Google probably knows more about me than 23andMe does, it’s pretty creepy. They help me find everything that I need: recipes, academic papers, anything. Obscure things, common things, helps me spell things, and in turn it knows all of my everything. Except for my genome, I guess. Maybe it has inferred huge chunks of it ... What other personal information do they have? Everything. They know where I live, they know my credit card number, they know what kind of hair I have, they know everything. I can’t think of anything that they don’t know honestly except for yeah, my DNA, I guess. You have to really dig down deep to find something that Google doesn’t know about me.” [P14] In sum, the participants unambiguously viewed their qualitative search queries as a far greater risk to their information privacy than their DNA because they thought their search queries revealed more about lives—their thoughts, feelings, concerns, and behavior—than their DNA possibly could.

While 23andMe users provided the company with a sample that is uniquely and immutably identifiable, the participants generally considered their DNA neither personal nor highly sensitive. From the perspective of a privacy researcher, on its face this conclusion appears incongruous. Your DNA will identify you forever, and potentially your relatives and offspring as well. What could pose a greater risk to one’s privacy, if a key attribute of privacy is one’s ability to remain unidentified, or even anonymous?

First, it illustrates the extent to which the participants’ existing privacy concerns are affected by temporality, tangibility, and salience to the context at hand. Most understood through experience that their search engine collected their queries and used them for targeted ads and other forms of behavioral analysis. There was no precedent for applying the same knowledge to their DNA; more abstract and forward-thinking risks were simply not on their radar, especially absent any recent scandals such as the Golden State Killer case. Next, it highlights the difficulty of acknowledging and accounting for collective, or networked, privacy effects in these systems, in particular when it is not in the company’s financial interest to do so. The core appeal of DTCGT is the individual payoff; to the extent that customers are encouraged to think about any collective effects of sharing their data, the sharing is framed in terms of contributing to a collective good via the indirect generalized exchange.

On one level, this disconnect between an expert view of the risks posed by DTCGT and the participants’ perceptions may primarily be that of experience. Given the relative novelty of DTCGT and the fact that the privacy risks are speculative and forward-thinking, it is no surprise that most consumers would be largely unaware of them. This might especially be true with DTCGT early adopters, though clearly more research examining the differences between those willing to engage with DTCGT and those who are not is needed. The fact that the privacy concerns that participants did have were based on contextual integrity—focusing on expectations of anonymity and reasonable uses of their data—also supports that conclusion. The participants’ concerns were

focused on being the *object* of inferences and, in turn, being *targeted* by those who make those inferences, particularly when the inferences were out of scope from the original intent of sharing their data. At the same time, this is not to say that anonymity was unimportant to them. In fact, the participants had a nearly universal expectation that the company would respect the contextual integrity of their genetic data by using it anonymously, particularly in research contexts.⁶ This expectation of anonymity was built upon the assumption that companies find the primary value of these databases in their aggregation of individuals, and that any individual contribution each user makes is miniscule.

The participants also relied upon a layer of *abstraction*, an expectation that no individualized surveillance was occurring, at least by other people, and that dehumanized algorithms were responsible for any inferences made by these companies. Several participants in fact expressed confidence that they believed computers, not people, were analyzing them. Whether the participants expected true anonymity—meaning, that the data used for analysis was scrubbed of any personally identifiable information (if that is even possible with DNA)—or practical obscurity, meaning that the transaction costs of identifying them among millions of other users provided them with *de facto* anonymity—is beyond the scope of this analysis. But what is clear is that the participants did not expect to be identified and targeted on an individual basis. These expectations also acted as an assurance to the participants in that their assumptions of anonymity and aggregate analysis minimized the risks of disclosure. And, because there was no obvious evidence at the time of the interviews that 23andMe was engaging in any practices that clearly violated the participants' contextual expectations, they generally felt their experience matched their expectations.

These findings have complex implications for information privacy and design. First, they help explain behaviors that some characterize as paradoxical: *why continue to use a service that gathers your personal information and violates your privacy?*, so goes the reasoning behind the privacy paradox. Assuming one has real alternatives available to them and could choose a different service (which isn't always the case), to the extent that individuals feel that they can hide in the database and that companies aren't interested in them as individuals provides some rationalization for the continued use of these services. Yes, the company knows who I am, but they are interested in me as a demographic category, not as an individual. This may have been true in the past, but as individual tracking has evolved to become a cross-platform and highly sophisticated operation, companies know more about us as individuals than ever before.

A few of these themes and their impact on disclosure merit more discussion, specifically: the impact of social motivations, assumptions of obscurity, the intangibility of future risks, the impact of context, and the consequences of networked data. Each will be discussed in turn below.

5.1 Social Motivations and Obscurity

Appealing to our social motivations—*we're all in this together*, or in the specific case of 23andMe, *"Becom[e] part of something bigger"*—focuses on the appeal of contributing to the company's research efforts and shifts the focus from how each individual actor directly benefits to how one's contribution may yield indirect benefits to both the individual contributor and the collective group. The focus on collective participation also offers assurances about privacy when, again, you're one small drop in a large ocean of data. But it is a false assurance; the company may actually know a great deal about you, but keeping this fact obscured allows individuals to believe in the illusion that their privacy is still intact. To the extent that companies can frame their operations this way

⁶23andMe states that it strips all identifying information from customer DNA before including it in research analysis unless otherwise specified. See <https://www.23andme.com/about/tos/>.

allows them to continue to collect data without restraint while giving both themselves and us a level of deniability.

This paper is not suggesting that 23andMe's research efforts will not yield any public benefit. But with the exception of some research efforts that require a specific genetic mutation for participation (e.g., lupus or Parkinson's), one's research participation occurs at a macro level, and individuals cannot electively choose in which public or private effort to participate. Notably, the framing of this appeal matters; for example, the company's website highlights the benefits of their research to public health, but does not explain how the company benefits.[25] Nor do they highlight in any detail their partnerships with private companies that don't focus on public health. 23andMe isn't alone in this approach. FamilyTreeDNA, the company that found themselves under scrutiny for allowing law enforcement open access to their database without informing their customers, cravenly reframed this access in April 2019 by renaming it "crowdsourcing," encouraging genetic testing users to upload their DNA to the site in order to help solve crimes, in essence, by exposing their family members—and violating their original customers' contextual integrity.[19] It seems reasonable that similar framing might be utilized to stimulate contributions to consumer ancestral databases as well, to the extent that mapping human relationships is viewed as a public benefit; a key differentiator in distinguishing these efforts from pure surveillance capitalism where the customers are the product, as technology critic Shoshanna Zuboff describes it, will be the answers to the questions of: *Who knows? Who decides? Who decides who decides?*[68]

5.2 Intangibility of Future Risk

Next, these findings reveal how difficult it is for people to factor in issues of future risk, particularly when those risks are intangible and not immediately salient. Because direct to consumer genetic testing (DTCGT) is arguably still in its infancy even in 2019, it seems reasonable to assume that the participants would have limited knowledge of the future-focused privacy risks DTCGT poses, particularly those with long-term implications. As Alessandro Acquisti and his colleagues have demonstrated, privacy decisions are especially vulnerable to temporal discounting as well as the intangibility of potential risks or consequences.[5][7] DTCGT is no exception; given that many of the risks are intangible or forward-thinking rather than immediate, it is no surprise that the participants in this study were unaware of them, or found them highly unlikely to affect them.

5.3 The Impact of Context

In addition, one's level of privacy concern is affected both by the immediate context and the salience of risk to that context.[44] If the risks appear distant and there are no contextual cues indicating danger, privacy concerns are unlikely to be salient at the time one makes a disclosure choice. As discussed, not only did these participants willingly seek out this service, the contextual cues they encountered provided them with assurance that the risks to using 23andMe were minimal to none. It was after this study was conducted that the news about the Golden State Killer broke. Suddenly, the reality that law enforcement could potentially include you in a search for criminal suspects without cause made it clearer that genetic data was a resource of immense value outside of the immediate commercial context. Multiple news stories featured quotes from angered DTCGT customers concerned with the violation of context: their genetic data, submitted for one express purpose, was now being used for another, without their consent. [31] Similar contextual concerns were raised when 23andMe partnered with GlaxoSmithKline and also sold access to their database of Parkinson's affected users to Genetech, but with far less public traction.[8]

Despite 23andMe not being an actual medical practitioner, the participants' primary association of DTCGT was health-related. Combined with the triviality of providing the sample itself (saliva), this context normalized the process of providing one's DNA to a DTCGT company. In short, there is

little about the process and not enough widely known about the potential consequences that would trigger most people to consider DTCGT unusually risky. Absent another source of information to the contrary, it is unsurprising that the participants either did not identify any salient risks, or if they did, found them to be abstract and unlikely at best.

While there is a compelling argument to be made for the need of widespread participation in DNA analysis for public health, 23andMe is a private company that makes no explicit guarantee that its data will be used to benefit either its contributors or the public broadly. However, it is the intangibility of these risks, and our human limitations to both anticipate long term consequences and engage in both temporal and hyperbolic discounting that make it difficult for anyone, including privacy experts, to predict how these issues will ultimately play out.^[6] But it is the strength of the assurances present today in the exchange relationships with 23andMe that mitigate enough of this uncertainty for participants to proceed with confidence.

5.4 Individual Versus Collective Privacy

One of the thorniest issues these findings raise is that by limiting our thinking and understanding of privacy to an individual problem, we lose sight of the fact that there can be impacts on privacy that affect individuals through their networks, presenting risk to the network itself. DTCGT highlights the inherent tension in taking individual personal information and revealing the connections in common with others: the balance between individual choices versus how those choices affect the larger community. This tension manifests itself in two ways on 23andMe's platform: through relative matching, and contributions to the company's research platform. Both of these mechanisms implicate the privacy of individuals who may not be participants on the service but whose privacy can nonetheless be affected. Further, DNA is inherently about human connections; it is a form of personal data that both pinpoints what is unique to us as individuals, as well as illuminating what we share in common with others, from our closest genetic family members to all of humanity through the ages. In isolation it may compromise our individual privacy by revealing the vulnerabilities in our inherent, biological selves. When connected, it provides the tools for building a relational graph based on genetics rather than social associations, which could be used for good or ill.

How to reconcile this tension between the choices of individuals and their potential threats to others? For example, I may have a legitimately compelling need to find out more about my genetic relationships today, a need I feel presently trumps the future privacy needs of my children. How do we weigh whose needs outweighs whose concerns? Marwick and boyd suggest a way forward: "[a] theory of networked privacy suggests that we must re-conceptualize the harms of privacy. Rather than thinking about privacy harms in terms of individuals or groups—classes of people—we need to frame privacy in terms of networks, or the relationships between people."^[36]

In terms of DTCGT, this suggests a duty to look beyond individual privacy concerns to the risks that arise to both subsets of the network as well as the network itself. What are the implications for an entire ethnic group, such as Uighur Muslims, to be mapped and identified through their DNA? What are the benefits and risks to all of humanity to be known in this way? These are difficult questions to grapple with, and unfortunately this paper cannot presently answer them. But in terms of implications for design, it suggests that individuals cannot be the only unit of analysis. This is true not just for questions of privacy specifically, but also for ethics more generally. There are risks and harms that simply won't reveal themselves when we examine these issues only from the perspective of the individual.

5.5 Generalizability and Limitations

With regards to generalizing this sample to all DTCGT customers, or even all 23andMe customers, there are several limitations. First, it is a small convenience sample of ten 23andMe customers all

located in the same metropolitan region. Thus, there may be important differences in customer motivations between 23andMe and other services, as well as geographical and cultural differences in attitudes. Other factors important to understanding the privacy issues raised by DTCGT are the novelty of the testing itself and the public's limited awareness of its risks, such that since these interviews were conducted, there have been numerous high profile criminal cases investigated using DTCGT companies' databases of customers without these customers' express consent. These cases may have since raised greater awareness among these participants of the possibility of risks that weren't widely discussed at the time of the interviews. Thus, both their privacy concerns and perceptions of risk, as well as the public in general, may have shifted at least slightly and likely will continue to shift as the tested population grows as long as similar incidents continue to occur.

Next, it is important to emphasize that this is a sample comprised of *willing* DTCGT users. Samples of people unwilling to take consumer genetic tests, as well as those who are undecided about doing so, may yield different results with regards to privacy concerns and knowledge of potential risks. For example, Horowitz *et al* conducted a survey in 2015 of nearly 110,000 U.S. adults registered with the National Marrow Donor Program, where they asked respondents if they would be willing to take a hypothetical free genetic ancestry test.[27]⁷ The authors report that 2% of their sample was not interested in taking a free ancestral genetic test, providing a range of reasons why, including privacy concerns, though general disinterest was the most cited reason. The factors significantly associated with disinterest in ancestral genetic testing were being a first-generation immigrant, as well as identifying as Asian. Participants who reported high confidence in their ancestral knowledge were also not interested.[27] More extensively examining the privacy concerns and perception of risk among people uninterested or unwilling to take these tests is an area that would benefit from future research.

6 IMPLICATIONS FOR DESIGN

These findings suggest implications for design that impact the application of theory and policy more than specific design choices. First, the study demonstrates the utility of using social exchange theory as a method for exploring relationships based on personal disclosure. In particular, SET is useful for identifying the structural aspects of institutional disclosure relationships that could otherwise be overlooked if one drew only from interpersonal privacy theories. For example, the participants in this study felt strongly that the direct negotiated exchange relationships they had with 23andMe—an exchange of DNA and money for a personalized genetic analysis—was fair to them and provided them a clear benefit. Additionally, possessing some amount of negotiative power in the form of control over their individual disclosure, specifically needing to opt-in to any additional information sharing, provided participants with a sense of agency and mitigated perceptions of power imbalances between the participants and the company. These factors, and the fact that the availability of other DTCGT services remains competitive (constraining any single company from enacting more onerous terms on its customers), illustrate the complexity of structural factors that weigh into personal disclosure decisions. From this perspective, one's decision to disclose her DNA to a DTCGT company isn't irrational or paradoxical to privacy. In fact, it is consonant with the findings of Brandimarte, Acquisti, and Loewenstein's study of privacy and personal disclosure, where they identified a "paradox of control": that as individual control over disclosure increases, one's willingness to disclose also increases, and that lower individual control results in less disclosure of personal information.[12] The paradox of control appears especially relevant when considering the participants' lack of awareness of privacy risks, their misperceptions

⁷This study focused exclusively on ancestral tests, not inherited traits and diseases, unlike 23andMe and other companies offering hybrid products.

about the 'medical' context and applicable laws in which these services exist; and the framing of the collection of DNA as something trivial and potentially worthless ("It's just spit").

At the same time, it demonstrates the challenges in studying information privacy from an individual's perspective when the risks are beyond an individual's control. For example, the fact that the respondents had not as of yet experienced a privacy violation by the company as conceptualized by contextual integrity does not mean that no privacy risks were present; it is that many of the privacy risks are inherent to the state of individual data being connected to others' data. As such, they are often too abstract for many individuals to grasp without some signal of a clear and present danger. But why would DTCGT companies signal such risks when doing so may disincentivize potential customers from signing up?

Due to the unique property of DNA as a biometric identifier in the U.S., regulation is needed in this space beyond GINA to protect both individuals and society. At a minimum, companies should voluntarily adhere to a set of principles such as those recommended by the Future of Privacy Forum, which set some basic ground rules for the industry.[21] These also should include the secondary services that will take your sequenced DNA and use it to provide services purportedly based on your unique genetic attributes. But as a matter of principle, individuals shouldn't have to make trade-offs between the very real benefits that DTCGT has brought many people, and fundamental risks to their genetic privacy. There are multiple culprits here: governments for being slow to regulate categories of personal information that are highly sensitive; the medical research establishment for overweighting the promise of collecting genetic data for public health while attending less to probable harms; and the companies themselves for architecting systems that leaves individuals' genetic data exposed rather than employing techniques such as differential privacy to make the samples themselves unidentifiable.[60]

There are a few discrete solutions that designers faced with working within spaces with either emergent or collective privacy concerns should consider. One is to fully question the existing mechanisms by which we provide notice and seek consent. Not only are they inadequate when the stakes are low, when the stakes are high it becomes far more urgent to engage with users fairly and honestly about possible risks. This failure is more urgent when the data under consideration is biometric. As Phillips notes, the "traditional 'notice and consent' model is insufficient" in the DTCGT field.[36] She suggests a higher standard of consent is necessary in DTCGT, due to consent requirements in the European Union's General Data Protection Regulation, as well as the nature of genetic information itself. Whether DTCGT should be subject to the medical standard of informed consent should be considered. Given that the risks of DTCGT are poorly known or understood by the public, and that the topic itself is inherently complex, from a consumer protection standpoint it is irresponsible to not require that consumers are provided both education and facts about DTCGT prior to taking a test.

Another is to grapple directly with the ethics of creating products or services that pose a threat to collective privacy, and to think through options and solutions for users that would both inform them of the risk, allow them to mitigate it, or withdraw entirely. We must note, for example, that one *can* use 23andMe specifically without electing to participate in either the relative matching or research platforms. You can request that they destroy your physical sample (if you consented to having it stored for future analysis), and after obtaining your testing results request that your account be deleted. However, the company encourages participation with these features, and notes in the help pages that they will not permanently delete your genetic information due to legal obligations, a fact that is only obliquely disclosed in their privacy policy.⁸ This creates a "Roach

⁸See <https://customercare.23andme.com/hc/en-us/articles/212170688-Requesting-account-closure>: "23andMe and our contracted genotyping laboratory will retain your Genetic Information, Date of Birth, and sex as required for compliance with

Motel" model of personal data where once your data checks into the company, it never checks out.[1] Legal compliance issues aside, this has terrible implications for user privacy and personal autonomy over one's data. To return to Marwick and boyd, co-construction, and/or participatory design, would be vital for ensuring that the intersections of context, individual privacy concerns, and collective privacy issues are addressed.

6.1 Directions for Future Work

This research was focused on understanding users' perceptions of privacy risk and expectations in an emergent area, with an emphasis on institutional privacy risks rather than social risks. However, the topic is ripe for extensive ethnographic research on all aspects of DTCGT, including an extensive exploration of the social turbulence created through the unexpected identification of familial ties. Further, issues of ethnic and racial identification through genetic testing is an area that deserves more scrutiny, especially given the arbitrary nature of how these platforms define these concepts. Additional inquiries into privacy and risk as the field grows should continue.

7 CONCLUSION

DTCGT promises its customers a novel way to sort through the puzzle pieces of their selves, through discovering one's inborn traits, suggesting one's racial and ethnic lineage, and finding unknown familial connections. It also offers the potential to contribute to paradigm shifting scientific research that promises no less than the discovery of breakthrough cures for inherited diseases, mapping the world's cultural heritage, and making significant contributions to our collective knowledge about humanity. The field is new, the promises are exciting, and the barrier to participation minimal for many.

The challenge to this optimism arises from questioning who is doing the collecting, who is accumulating the knowledge, and who it benefits most. While the participants in this study had no illusions that they were engaging with a private company that would profit from their genomic contributions, they nonetheless lacked a general awareness of potential risks and held an uncritical belief in the benefits of the company's research efforts. They assumed that anonymity would ultimately protect them from potential privacy harm, and that their ultimate contribution was being one small part of a large and more valuable whole.

DTCGT in its own way presents a privacy paradox. Like many services offered at scale, its reliability and usefulness for everyone increases as more people contribute to it. And yet, DTCGT services offer a unique privacy threat not only for individual customers, but for society at large. Not only do these services collect a singular form of data which identifies us all uniquely for perpetuity, they also inherently expose the genetic relationships between all of us, making it probable that most humans will lose their ability to be both unknown and unidentifiable. While admittedly one's ability to exist outside either public or commercial structures is already challenged in digitally connected societies, DTCGT could be the final nail in the coffin of anonymity, as even those people who don't have a traceable digital presence find themselves drawn into a genetic human map through their relatives.

While privacy scholars have documented the many ways in which people construct and navigate social privacy in digital spaces (see generally:[42][47][35][34][65]) and suggested design interventions, the collective identification presented by DTCGT takes the challenge of managing collective privacy to a new level while also raising critical questions of consent. The implication of sharing of one's data held in common with one's genetic relatives is similar to the data sharing enabled

legal obligations, pursuant to the California laboratory regulations and the US. Federal Clinical Laboratory Improvement Amendments of 1988 (CLIA), respectively."

by Facebook's now infamous friends permission (which enabled Cambridge Analytica to collect the profile information of over 80 million Facebook users without their consent), where decisions by a relative few exposed the information of many of millions more. And yet, if the responses by the participants in this study can be generalized, most would likely be more upset with the dissemination of their Facebook profile data than their DNA, given that the qualitative nature of Facebook data can reveal one's emotions, likes, dislikes, and concerns: the lived experience of being human.

In the absence of legislation limiting the collection or usage of DNA in the consumer context, as well as the dissemination of information about potential risks, DTCGT companies and their secondary market partners are creating a mythology around genetic testing based on the triviality of the physical contribution ("It's just spit"), promises of revealing hidden family histories and relationships, and appeals to the greater good of humanity through the generation of scientific knowledge acquired by sharing's one DNA with researchers. The challenge to these optimistic promises is that they rely on private actors consistently making decisions that benefit their customers over their own interests. For example, all of these various actors must practice good security and ensure your DNA isn't stolen by hackers, or is compromised by a third party with whom they chose to share data. They must be willing and able to protect their customers' data against incursions by governmental authorities around the globe. Unfortunately, given the track record of most data collectors, this requires more than just optimism; it requires blind faith.

ACKNOWLEDGMENTS

The author would like to thank: Rena Coen, Qing Huang, and Matthew Nagamine for their research assistance; Deirdre Mulligan and Coye Cheshire for their invaluable support; Andreas Katsanevas and Jana Gooth for their crucial feedback; and the anonymous reviewers of this paper, whose time and expertise improved the final paper.

This research was supported by grants from the Center for Long-Term Cybersecurity at U.C. Berkeley and the Center for Technology, Society, and Policy at the U.C. Berkeley School of Information.

REFERENCES

- [1] [n. d.]. https://en.wikipedia.org/wiki/Roach_Motel
- [2] 23andMe. [n. d.]. 23andMe - Research.
- [3] 23andMe. [n. d.]. 23andMe.com - How it Works.
- [4] 23andMe. [n. d.]. 23andMe.com Terms of Service.
- [5] A Acquisti, L Brandimarte, and G Loewenstein. 2015. Privacy and human behavior in the age of information. *Science* 347, 6221 (2015), 509–514.
- [6] Alessandro Acquisti and Jens Grossklags. 2007. What can behavioral economics teach us about privacy? In *Digital Privacy: Theory, Technologies, and Practices*, A. Acquisti, S. Gritzalis, S. Di Vimercati, and C. Lambrinoudakis (Eds.). Auerbach Publications, 363–379.
- [7] Alessandro Acquisti, Curtis R Taylor, and Liad Wagman. 2015. The Economics of Privacy. *SSRN Electronic Journal* (2015). https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2580411
- [8] Shelin Adam and Jan M. Friedman. 2016. Individual DNA samples and health information sold by 23andMe. *Genetics in Medicine* 18, 4 (apr 2016), 305–306. <https://doi.org/10.1038/gim.2015.82>
- [9] Irwin Altman. 1975. *The Environment and Social Behavior*. Brooks/Cole Publishers.
- [10] Eli Badan-Lasar. 2019. A Family Portrait: Brothers, Sisters, Strangers. *The New York Times* (June 2019). <https://www.nytimes.com/interactive/2019/06/26/magazine/sperm-donor-siblings.html>
- [11] Louise Barkhuus. 2012. The Mismeasurement of Privacy: Using Contextual Integrity to Reconsider Privacy in HCI. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '12)*. ACM, New York, NY, USA, 367–376. <https://doi.org/10.1145/2207676.2207727>
- [12] Laura Brandimarte, Alessandro Acquisti, and George Loewenstein. 2012. Misplaced Confidences: Privacy and the Control Paradox. *Soc. Psychol. Personal. Sci.* 4, 3 (May 2012), 340–347. <https://doi.org/10.1177/1948550612455931>

- [13] Alex Braunstein, Laura Granka, and Jessica Staddon. 2011. Indirect Content Privacy Surveys: Measuring Privacy Without Asking About It. In *Proceedings of the Seventh Symposium on Usable Privacy and Security (SOUPS '11)*. ACM, New York, NY, USA, Article 15, 14 pages. <https://doi.org/10.1145/2078827.2078847>
- [14] Yunan Chen and Heng Xu. 2013. Privacy Management in Dynamic Groups: Understanding Information Privacy in Medical Practices. In *Proceedings of the 2013 Conference on Computer Supported Cooperative Work (CSCW '13)*. ACM, New York, NY, USA, 541–552. <https://doi.org/10.1145/2441776.2441837>
- [15] Coye Cheshire, Alexandra Gerbasi, and Karen S. Cook. 2010. Trust and Transitions in Modes of Exchange. *Social Psychology Quarterly* 73, 2 (2010), 176–195.
- [16] Hichang Cho and Anna Filippova. 2016. Networked Privacy Management in Facebook: A Mixed-Methods and Multinational Study. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing (CSCW '16)*. ACM, New York, NY, USA, 503–514. <https://doi.org/10.1145/2818048.2819996>
- [17] Susan Dominus. 2019. Sperm Donors Can't Stay Secret Anymore. Here's What That Means. *The New York Times* (June 2019). <https://www.nytimes.com/2019/06/26/magazine/sperm-donor-questions.html>
- [18] Jamie Ducharme. 2018. A Major Drug Company Now Has Access to 23andMe's Genetic Data. Should You Be Concerned? *Time Magazine* (July 2018). <https://time.com/5349896/23andme-glaxo-smith-kline/>
- [19] Sarah Emerson. 2019. A DNA Ancestry Website Is Asking People to Join to Help Hunt Criminals. *Motherboard - Vice.com* (March 2019). https://motherboard.vice.com/en_us/article/bjqebq/a-dna-ancestry-website-is-asking-people-to-join-to-help-hunt-criminals?utm_source=mbtwitter
- [20] Yaniv Erlich, Tal Shor, Itsik Pe'er, and Shai Carmi. 2018. Identity inference of genomic data using long-range familial searches. *Science (New York, N.Y.)* 362, 6415 (nov 2018), 690–694. <https://doi.org/10.1126/science.aau4832>
- [21] Future of Privacy Forum. 2018. *Privacy Best Practices for Consumer Genetic Testing Services*. Technical Report. Future of Privacy Forum. <https://fpf.org/wp-content/uploads/2018/07/Privacy-Best-Practices-for-Consumer-Genetic-Testing-Services-FINAL.pdf>
- [22] Thomas Fuller. 2018. How a Genealogy Site Led to the Front Door of the Golden State Killer Suspect. *The New York Times* (April 2018). <https://www.nytimes.com/2018/04/26/us/golden-state-killer.html?module=inline>
- [23] Bob Gellman. 2017. FAIR INFORMATION PRACTICES: A Basic History. <https://bobgellman.com/rg-docs/rg-FIPshistory.pdf>
- [24] Taylor Hatmaker. 2018. DNA analysis site that led to the Golden State Killer issues a privacy warning to users. *Techcrunch.com* (May 2018).
- [25] Matthew Herper. 2015. Surprise! With \$60 Million Genentech Deal, 23andMe Has A Business Plan. *Forbes* (January 2015). <https://www.forbes.com/sites/matthewherper/2015/01/06/surprise-with-60-million-genentech-deal-23andme-has-a-business-plan/>
- [26] Tina Hesman Saey. 2019. What FamilyTreeDNA sharing genetic data with police means for you. *Science News* (February 2019).
- [27] Adam L. Horowitz, Aliya Saperstein, Jasmine Little, Martin Maiers, and Jill A. Hollenbach. 2019. Consumer (dis-)interest in genetic ancestry testing: the roles of race, immigration, and ancestral certainty. *New Genetics and Society* (jan 2019), 1–30. <https://doi.org/10.1080/14636778.2018.1562327>
- [28] Hsiao-Ying Huang and Masooda Bashir. 2015. Direct-to-consumer Genetic Testing: Contextual Privacy Predicament. In *Proceedings of the 78th ASIS&T Annual Meeting: Information Science with Impact: Research in and for the Community (ASIST '15)*. American Society for Information Science, Silver Springs, MD, USA, 50:1—50:10. <http://dl.acm.org/citation.cfm?id=2857070.2857120>
- [29] International Genome Sample Resource. [n. d.]. IGSR: The International Genome Sample Resource. <http://www.internationalgenome.org/about>
- [30] Brendan I Koerner. 2015. Your Relative's DNA Could Turn You Into a Suspect. *WIRED* (October 2015). <https://www.wired.com/2015/10/familial-dna-evidence-turns-innocent-people-into-crime-suspects/>
- [31] Gina Kolata and Heather Murphy. 2018. The Golden State Killer Is Tracked Through a Thicket of DNA, and Experts Shudder. *The New York Times* (April 2018). <https://www.nytimes.com/2018/04/27/health/dna-privacy-golden-state-killer-genealogy.html?>
- [32] Airi Lampinen and Coye Cheshire. 2016. Hosting via Airbnb. In *Proceedings of the 2016 {CHI} Conference on Human Factors in Computing Systems - {CHI} '16*.
- [33] Airi Lampinen, Vilma Lehtinen, Coye Cheshire, and Emmi Suhonen. 2013. Indebtedness and Reciprocity in Local Online Exchange. In *Proceedings of the 2013 Conference on Computer Supported Cooperative Work (CSCW '13)*. ACM, New York, NY, USA, 661–672. <https://doi.org/10.1145/2441776.2441850>
- [34] Airi Lampinen, Vilma Lehtinen, Asko Lehmuskallio, and Sakari Tamminen. 2011. We're in It Together: Interpersonal Management of Disclosure in Social Network Services. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11)*. ACM, New York, NY, USA, 3217–3226. <https://doi.org/10.1145/1978942.1979420>

- [35] Eden Litt and Eszter Hargittai. 2014. A Bumpy Ride on the Information Superhighway. *Comput. Hum. Behav.* 36, C (jul 2014), 520–529. <https://doi.org/10.1016/j.chb.2014.04.027>
- [36] Alice E Marwick and Danah boyd. 2014. Networked privacy: How teenagers negotiate context in social media. *New Media & Society* 16, 7 (2014), 1051–1067. <https://doi.org/10.1177/1461444814543995>
- [37] Scott McGrath and Dhundy (Kiran) Bastola. 2014. DTC Genetic Testing and Consumer Comprehension. In *Proceedings of the 5th ACM Conference on Bioinformatics, Computational Biology, and Health Informatics (BCB '14)*. ACM, New York, NY, USA, 582–583. <https://doi.org/10.1145/2649387.2660559>
- [38] Linda D. Molm and Karen S. Cook. 1995. Social Exchange and Exchange Networks. In *Sociological Perspectives on Social Psychology*, J. Fine, G. and House (Ed.). Allyn and Bacon, Needham, MA., Chapter 8, 209–235.
- [39] Linda D. Molm, Nobuyuki Takahashi, and Gretchen Peterson. 2000. Risk and Trust in Social Exchange: An Experimental Test of a Classical Proposition. *Amer. J. Sociology* 105, 5 (2000), 1396–1427. <https://doi.org/10.1086/210434>
- [40] National Geographic. [n. d.]. The Genographic Project. <https://genographic.nationalgeographic.com/about/>
- [41] Helen Nissenbaum. 2009. *Privacy in Context: Technology, Policy, and the Integrity of Social Life*. Stanford University Press, Stanford, CA.
- [42] Leysia Palen and Paul Dourish. 2003. Unpacking "Privacy" for a Networked World. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '03)*. ACM, New York, NY, USA, 129–136. <https://doi.org/10.1145/642611.642635>
- [43] Sandra S. Petronio. 2002. *Boundaries of Privacy: Dialectics of Disclosure*. State University of New York Press, Albany. 268 pages.
- [44] Chanda Phelan, Cliff Lampe, and Paul Resnick. 2016. It's Creepy, But it Doesn't Bother Me. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems - CHI '16*. <https://doi.org/10.1145/2858036.2858381>
- [45] Andelka M Phillips. 2016. Take an online DNA test and you could be revealing far more than you realise. *The Conversation* (2016). <http://theconversation.com/take-an-online-dna-test-and-you-could-be-revealing-far-more-than-you-realise-52734>
- [46] Andrew Pollack. 2015. 23andMe Will Resume Giving Users Health Data. *The New York Times* (October 2015). <https://www.nytimes.com/2015/10/21/business/23andme-will-resume-giving-users-health-data.html?>
- [47] Yu Pu and Jens Grossklags. 2017. Valuating Friends' Privacy: Does Anonymity of Sharing Personal Data Matter?. In *Proceedings of the Thirteenth USENIX Conference on Usable Privacy and Security (SOUPS'17)*. USENIX Association, Berkeley, CA, USA, 339–355. <http://dl.acm.org/citation.cfm?id=3235924.3235952>
- [48] Antonio Regalado. 2018. 2017 was the year consumer DNA testing blew up. *MIT Technology Review* (February 2018). <https://www.technologyreview.com/s/610233/2017-was-the-year-consumer-dna-testing-blew-up/>
- [49] Michelle Robertson. 2017. After a 23andMe DNA test, Northern California woman discovers 3 siblings. *San Francisco Chronicle* (September 2017). <http://www.sfgate.com/bayarea/article/23andme-genetic-dna-test-siblings-sperm-donor-12209687.php>
- [50] Wendy D. Roth and Katherine A. Lyon. 2018. Genetic ancestry tests and race: Who takes them, why, and how do they affect racial identities? In *Reconsidering Race: Social Science Perspectives on Racial Categories in the Age of Genomics*. Vol. 1. Oxford University Press, 133–169. <https://doi.org/10.1093/oso/9780190465285.003.0008>
- [51] Dan Ryan. 2006. Getting the Word Out: Notes on the Social Organization of Notification. *Sociological Theory* 24, 3 (2006), 228–254.
- [52] Johnny Saldaña. 2016. *The Coding Manual for Qualitative Researchers* (3rd ed.). Sage Publications. 368 pages.
- [53] Charles Seife. 2013. 23andMe Is Terrifying, but Not for the Reasons the FDA Thinks. *Scientific American* (November 2013). <https://www.scientificamerican.com/article/23andme-is-terrifying-but-not-for-the-reasons-the-fda-thinks/>
- [54] Pan Shi, Heng Xu, and Yunan Chen. 2013. Using Contextual Integrity to Examine Interpersonal Information Boundary on Social Network Sites. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13)*. ACM, New York, NY, USA, 35–38. <https://doi.org/10.1145/2470654.2470660>
- [55] Frederic Stutzman and Woodrow Hartzog. 2012. Boundary Regulation in Social Media. In *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work (CSCW '12)*. ACM, New York, NY, USA, 769–778. <https://doi.org/10.1145/2145204.2145320>
- [56] Jennifer Jiyoung Suh, Miriam J Metzger, Scott A Reid, and Amr El Abbadi. 2018. Distinguishing Group Privacy From Personal Privacy: The Effect of Group Inference Technologies on Privacy Perceptions and Behaviors. *Proc. ACM Hum.-Comput. Interact.* 2, CSCW (nov 2018), 168:1—168:22. <https://doi.org/10.1145/3274437>
- [57] Emmi Suhonen, Airi Lampinen, Coye Cheshire, and Judd Antin. 2010. Everyday Favors: A Case Study of a Local Online Gift Exchange System. In *Proceedings of the 16th ACM International Conference on Supporting Group Work (GROUP '10)*. ACM, New York, NY, USA, 11–20. <https://doi.org/10.1145/1880071.1880074>
- [58] U.S. Food and Drug Administration. 2017. FDA allows marketing of first direct-to-consumer tests that provide genetic risk information for certain conditions. <https://www.fda.gov/NewsEvents/Newsroom/PressAnnouncements/ucm551185.htm>

- [59] Na Wang, Pamela Wisniewski, Heng Xu, and Jens Grossklags. 2014. Designing the Default Privacy Settings for Facebook Applications. In *Proceedings of the Companion Publication of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing (CSCW Companion '14)*. ACM, New York, NY, USA, 249–252. <https://doi.org/10.1145/2556420.2556495>
- [60] Shuang Wang, Xiaoqian Jiang, Siddharth Singh, Rebecca Marmor, Luca Bonomi, Dov Fox, Michelle Dow, and Lucila Ohno-Machado. 2017. Genome privacy: challenges, technical approaches to mitigate risk, and ethical considerations in the United States. *Annals of the New York Academy of Sciences* 1387, 1 (2017), 73–83. <https://doi.org/10.1111/nyas.13259>
- [61] Sui-Lee Wee. 2019. China Uses DNA to Track Its People, With the Help of American Expertise. *The New York Times* (February 2019). <https://www.nytimes.com/2019/02/21/business/china-xinjiang-uighur-dna-thermo-fisher.html?>
- [62] Alan Westin. 1967. *Privacy and Freedom*. Atheneum, New York.
- [63] Benjamin Winterhalter. 2014. A genetic 'Minority Report': How corporate DNA testing could put us at risk. *Salon* (January 2014). https://www.salon.com/2014/01/26/a_genetic_minority_report_how_corporate_dna_testing_could_put_us_at_risk/
- [64] Pamela Wisniewski, A K M Najmul Islam, Bart P Knijnenburg, and Sameer Patil. 2015. Give Social Network Users the Privacy They Want. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing (CSCW '15)*. ACM, New York, NY, USA, 1427–1441. <https://doi.org/10.1145/2675133.2675256>
- [65] Pamela Wisniewski, Heather Lipford, and David Wilson. 2012. Fighting for My Space: Coping Mechanisms for Sns Boundary Regulation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '12)*. ACM, New York, NY, USA, 609–618. <https://doi.org/10.1145/2207676.2207761>
- [66] Sarah Zhang. 2017. The Loopholes in the Law Prohibiting Genetic Discrimination. *The Atlantic* (March 2017). www.theatlantic.com/health/archive/2017/03/genetic-discrimination-law-gina/519216/
- [67] Sarah Zhang. 2019. The Fertility Doctor's Secret. *The Atlantic* (March 2019). <https://www.theatlantic.com/magazine/archive/2019/04/fertility-doctor-donald-cline-secret-children/583249/>
- [68] Shoshanna Zuboff. 2019. *The Age of Surveillance Capitalism*. Public Affairs, New York, New York, USA. 691 pages.

Appendices

A GENETIC TESTING CONTEXT-SPECIFIC QUESTIONS

Interview Questions: Genetic Testing Context	
Lead Question	Lead Question: Let’s start first with how you came to use genetic testing, and why you wanted to have it done. Can you describe to me your decision-making process, and how you chose the service you used?
Follow-Up Questions	<ul style="list-style-type: none">• What made you decide to have genetic testing performed?• When did you remember first hearing about 23andMe?• What was your first impression?• What were your impressions based on?• What did their website make you feel about them?• What did you like/dislike about the service?• Have you ever had any other form of genetic testing?• Did you consider any alternatives, such as getting tested through your doctor?

B INTERVIEW QUESTIONS PROBING SET-BASED THEORETICAL CONCEPTS

Interview Questions: SET Theoretical Concepts

Exploring the Nature of the Exchange Relationship	<ul style="list-style-type: none"> • Describe for me the nature of your relationship with 23andMe. For example, what do they offer you, and what do you offer them? • What do you get out of the relationship?
Fairness and Benefits	<ul style="list-style-type: none"> • Would you have used the service if it were free? • Was the price fair for what you provided and received in return? • What benefit is the company getting? • Who benefits more from this relationship: you or them? Or do you feel it is equal? • Do you think of the company as a partner or a friend? How do they treat you?
Power and Negotiation	<ul style="list-style-type: none"> • How easy or hard would it be for you to stop the relationship? • Do you feel like equals? Or does one of you have more advantage/power? • Do you feel as if you know as much about them as they know about you? • Do you feel as if you have a personal commitment to the company? • Is this a relationship you want to continue with in the future? • Has anything changed in how you thought about the service after you started using it as compared to how you felt when you signed up/first started using?
Trust and Assurances	<ul style="list-style-type: none"> • What made you comfortable/trust sharing your data with the company? • Are there any other factors that made you comfortable using this company? • For example, is there anything specific that made you feel safe, or offered you protection or a guarantee? • If the company did something that you felt violated your "agreement" with them, angered you, or felt unfair or illegal, what would you do? Who would you complain to about the company? • Do you recall if you saw a privacy policy for the service? IF YES: did you read it before you signed up? Were there any aspects of the policy that informed your decision to use the service? • Do you know if there are any laws that protect the information you provided to the company?

C INTERVIEW QUESTIONS FOCUSING ON PRIVACY, RISK, AND DISCLOSURE

Interview Questions: Privacy, Risk, and Personal Disclosure

Personal Disclosure	<ul style="list-style-type: none">• What kind of personal information did you provide to the company?• Have you shared your information on the company’s platform?• How has your interaction with the company affected your expectations with the service?• Has there been anything unexpected in your interactions with company/app?• Is it a 1-to-1 relationship? Meaning, is information is going directly to company/app and not anywhere else?• Do you feel as if there are aspects of the service that you can opt out of or just not use? If yes, what aspects?• Do you think you can provide company with feedback or have them make a change that would better suit your needs?
Risks and Trade-Offs	<ul style="list-style-type: none">• Since you first signed up with the company, do you trust them more or less now, or the same?• What, if anything, would make you stop using this service?• Does how they handle your personal information meet your expectations?• Who else do you trust with the type of information you provide to this company?• Do you know of any settings or controls that enable you to manage how your information is used?• Are there are any downsides to using this service?• Have you shared your results with your doctor or other health professional?• Would you want your health insurance company to have your testing results?

Received April 2019; revised June 2019; accepted August 2019