Supporting the Contact Tracing Process with WiFi Location Data: Opportunities and Challenges

Kaely Hall
khall33@gatech.edu
Georgia Institute of Technology
Atlanta, Georgia, USA

Dong Whi Yoo
yoo@gatech.edu
Georgia Institute of Technology
Atlanta, Georgia, USA

Wenrui Zhang
zhangwenrui@gatech.edu
Georgia Institute of Technology
Atlanta, Georgia, USA

Mehrab Bin Morshed
mehrab.morshed@gatech.edu
Georgia Institute of Technology
Atlanta, Georgia, USA

Vedant Das Swain
vedantswain@gatech.edu
Georgia Institute of Technology
Atlanta, Georgia, USA

Gregory D. Abowd
g.abowd@northeastern.edu
Northeastern University
Boston, Massachusetts, USA

Munmun De Choudhury
munmun.choudhury@cc.gatech.edu
Georgia Institute of Technology
Atlanta, Georgia, USA

Alex Endert
endert@gatech.edu
Georgia Institute of Technology
Atlanta, Georgia, USA

John Stasko
john.stasko@cc.gatech.edu
Georgia Institute of Technology
Atlanta, Georgia, USA

Jennifer G Kim
jennifer.kim@cc.gatech.edu
Georgia Institute of Technology
Atlanta, Georgia, USA

ABSTRACT

Contact tracers assist in containing the spread of highly infectious diseases such as COVID-19 by engaging community members who receive a positive test result in order to identify close contacts. Many contact tracers rely on community member's recall for those identifications, and face limitations such as unreliable memory. To investigate how technology can alleviate this challenge, we developed a visualization tool using de-identified location data sensed from campus WiFi and provided it to contact tracers during mock contact tracing calls. While the visualization allowed contact tracers to find and address inconsistencies due to gaps in community member's memory, it also introduced inconsistencies such as false-positive and false-negative reports due to imperfect data, and information sharing hesitancy. We suggest design implications for technologies that can better highlight and inform contact tracers of potential areas of inconsistencies, and further present discussion on using imperfect data in decision making.

CCS CONCEPTS

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

KEYWORDS

contact tracing, WiFi data, visualization

ACM Reference Format:

Kaely Hall, Dong Whi Yoo, Wenrui Zhang, Mehrab Bin Morshed, Vedant Das Swain, Gregory D. Abowd, Munmun De Choudhury, Alex Endert, John Stasko, and Jennifer G Kim. 2022. Supporting the Contact Tracing Process with WiFi Location Data: Opportunities and Challenges. In *CHI Conference on Human Factors in Computing Systems (CHI '22), April 29-May 5, 2022, New Orleans, LA, USA.* ACM, New York, NY, USA, 14 pages. https://doi.org/10.1145/3491102.3517703

1 INTRODUCTION

Reducing the epidemic peak of infectious diseases such as COVID-19, or "flattening the curve," is essential to ensure that hospitals are not overwhelmed with patients, and that frontline workers and overall public health are not put in jeopardy [23, 36]. According to the Centers for Disease Control and Prevention (CDC), contact tracing is key to slowing the spread of infectious diseases [5]. Contact tracing is an example of a non-pharmaceutical intervention (NPI), and a vital first response to defend against a pandemic [36]. The benefit of NPIs is that they are the only set of pandemic countermeasures that are readily available at all times and in all countries, therefore making them the most accessible means of containment. These measures have proven to be effective both at the start and the during a pandemic outbreak [36]. Because new human pathogen species are discovered at a rate of 3 per year [33], contact tracing will continue to be required all over the world.

Contact tracing works to help slow the spread of infection by identifying and notifying community members of potential exposures and providing further information about quarantining and local resources [7]. However, unlike other infectious diseases such

CHI '22, April 29-May 5, 2022, New Orleans, LA, USA

ACM ISBN 978-1-4503-9157-3/22/04…$15.00
https://doi.org/10.1145/3491102.3517703

as Ebola and tuberculosis, COVID-19's unprecedented infection rate has presented the unique challenge of outpacing local public health systems' abilities to quickly identify and contact all who have been exposed. Another complicating factor of containment is due to 85% of COVID-19 transmission being from asymptomatic carriers [44]. This means that people may not be aware that they are positive for the virus due to not having symptoms, and are exposing others. This unique combination of factors require contact tracing efforts and resources to operate at peak efficiency [7].

The effectiveness of contact tracing depends on many factors, such as timeliness of case identification and contact notification, contact tracers' expertise, and self-reporting [15]. Multiple efforts have been undertaken to train contact tracers. For example, government health officials in California have trained over 200,000 contact tracers [32], and more than a million people have enrolled in an online contact tracing Coursera course offered by Johns Hopkins [4]. Manual contact tracing—when health workers directly communicate with an individual who has tested positive (i.e., a positive case) to determine each of their close contact contacts—relies on accurate and reliable identification of potential exposures [15]. However, because human recall can be unreliable, often contact tracers receive incomplete or incorrect information from the positive case [15, 55].

Digital support for contact tracing automatically monitors people's location and proximal interactions, handing off to technology the burden of determining a person's (or more specifically their on-person devices) history of proximity to other people. Many authorities and governing bodies, such as Harvard University and the Australian government, have adopted digital solutions to track potential close contacts of users [2, 6, 9, 35]. Many of these digital applications are intended to accelerate detection and notification of potential exposures [51]. However, little is known about how digital options can be integrated into the manual contact tracing process and support close contact identification. It is well documented that both manual and digital contact tracing data imperfections are present, however, more studies need to be conducted on how to successfully integrate digital tools into the process and how to address the challenge [15, 51].

To understand how digital contact tracing can effectively support the manual contact tracing process, we developed a visualization tool using community member's location information, gathered from University WiFi network. We situated our study in a university setting because they are dense population hubs for the community and therefore critical epicenters to control disease outbreak [14]. Furthermore, data collection via WiFi network was chosen because of the robust WiFi infrastructure employed on large University campuses, which can lead to high adoption rates without requiring extra software installation for each individual.

Our paper makes the following contributions:

- We present a novel visualization tool for manual contact tracing using WiFi location data. By employing the tool during mock contact tracing calls between 5 contact tracers and 14 community members, we show the way the visualization supports their conversation, and tensions its introduction can create.
- We report the finding of inconsistencies that are introduced into manual contact tracing as a result of using the visualization tool. We show how contact tracers and community members resolved the inconsistencies by engaging in collaborative sensemaking of the data.
- We present and discuss concerns within the community regarding privacy of using the tool visualizing WiFi location data for the purpose of contact tracing.
- We demonstrate the existence of socio-technical gaps within contact tracing technologies and suggest design implications to alleviate those gaps.

2 BACKGROUND: CONTACT TRACING FOR COVID-19

Contact tracing is cited by the CDC as a key component of controlling the transmission of infectious diseases such as COVID-19 [7]. Contact tracing involves the engagement of an infected person, or "positive case," by a public health worker. During this interaction, the public health worker, or "contact tracer," facilitates a discussion about their symptom timelines, test dates, and contacts [7]. In this section, we provide background of manual and digital contact tracing, as this study investigates how digital tools can be designed to better support the manual contact tracing process.

2.1 Manual Contact Tracing

In the context of pandemic response, a "contact" is anyone who came into physical contact with a positive case during their infectious period. More specifically, the CDC defines a "close contact" as anyone who has been within 6 feet (2 meters) of an infected person during their infectious period for 15 minutes or more over a 24 hours period [5]. In manual contact tracing, all contact is made via phone call, and most states in the U.S. currently adhere to manual contact tracing [32]. When the contact tracer calls the positive case, they ask a series of questions in order to gather information on infectious periods, events attended, and close contacts of the positive case. When the list of close contacts has been made, the contact tracer then calls each close contact to notify them of exposure and provide self-isolation timelines and information.

Because manual contact tracing relies entirely on the positive case's ability to provide information, imperfect recall presents a significant challenge to the process. [15]. There has been research done on hypothetical strategies to address the recall problem of manual contact tracing using fully digital contact tracing strategies [55], however, there needs to be more investigation of how digital strategies can be integrated to support manual contact tracing directly [51]. Our study sheds light on how the data collected by digital systems can be visualized and discussed during contact tracing calls can alleviate the recall problem that exists.

2.2 Technology Support for Contact Tracing

There are many ways that technology can be used to support contact tracing. Digital solutions can be used to create a list of potential close contacts, alert people of exposure, or both. For both functionalities, the digital system must be able to determine the proximity between two people. Below we will discuss two different ways technology can determine proximity between two people, and present characteristics of each that impact efficacy and adoption rates.

2.2.1 Models to Determine Proximity for Contact Tracing.

Proximity is defined as "the state of being near in space or time [8]." In the context of contact tracing, closer proximity to an infected person means higher risk of exposure [1]. There are multiple ways that technology can use to determine one's proximity to another. For contact tracing specifically, we explore two different models to determine proximity of community members: 1) decentralized device-based, and 2) centralized infrastructure-based.

Device-based proximity determination is based on the physical distance between two devices, and decentralized refers to the data being stored locally on the device [31]. This model utilizes radio communication capabilities such as Bluetooth to detect the strength of a signal coming from another device [55]. The strength of the signal is proportional to the distance between the devices at a particular time, and therefore determines proximity of the devices to each other. This interaction is called a "handshake," and this method is generally accurate and can be used to quickly alert user of potential exposure. In addition, because decentralized proximity determination is local to the device, it gives the user more control over the data and system use [28].

The second model explored to determine proximity is centralized infrastructure-based. Instead of tracking the distance between devices, infrastructure-based proximity determination tracks a device's location through a network, such as community WiFi [46], and the data is uploaded and stored in a centralized database. For example, when a mobile device comes into contact with a network via access point or other sensor, the devices' location data within the network is logged. If other users' devices are also interacting with the same network and sensor at the same time, the system can report that these two devices, and therefore their users, are in the same space at the same time. In this work, we refer to these users who are sharing space and time as "co-locators."

Another way centralized proximity determination can be used is by utilizing existing infrastructure such as C.C.T.V footage and credit card usage information, as South Korea is doing [25]. While users have less control over their data in centralized systems, they assess transmission risk more accurately [15]. Centralized proximity determination via WiFi has also been used in past studies to reconstruct the locations visited by people on campus [20, 50, 53].

2.2.2 Factors Influencing Adoption of Contact Tracing Technology.

In order to be effective in slowing infection rates of COVID-19, contact tracing applications require an adoption rate of at least 60% [31]. The adoption of contact tracing technology depends on several factors including privacy and security issues, benefits to users and their community, and accuracy [11, 26, 42, 51]. We review each factor below as they can inform potential benefits and challenges of introducing the centralized infrastructure-based digital contact tracing in manual contact tracing process.

Privacy: Decentralized and centralized approaches to logging user's proximity data have differing levels of data privacy and control [22]. Systems that store user proximity data via decentralized method typically preserve the privacy of the user, as the information is stored locally on their device [45]. The decentralized NOVID application is marketed for being designed explicitly around user privacy [21], as was Singapore's national contact tracing app TraceTogether, which uses Bluetooth to detect other devices [3].

Conversely, centralized systems log the user data in a network database. These systems, such as Tracefi and South Korea's controversial use of C.C.T.V. footage and credit card information [25], have raised privacy concerns around who has access to the data and what it can be used for [39]. People reported being less comfortable with the idea that their locations can be either made public or used for things other than contact tracing [39, 45]. However, these privacy concerns can arise with decentralized applications as well. Reports out of Singapore have stated that the local government officials are requesting access to the data in the decentralized TraceTogether application for purposes other than public health and pandemic response [52]. Furthermore, a recent study that was conducted specifically to understand people's attitudes toward centralized and decentralized applications for contact tracing revealed that people were more likely to trust a centralized approach, citing that they trusted a central authority more than a third party application [31]. This finding suggests that community members would likely adopt a privacy-preserving centralized system for contact tracing that is developed and maintained by a government or public health authority that they trust.

Public health benefit: People are willing to install digital contact tracing applications that can benefit public health such as reduction in infection rate [26]. In general, using a contact tracing technology can be viewed as a civic duty for the purpose of supporting public health efforts [55]. The centralized approach can provide benefits to public health as it can readily offer the number of people at risk because it already possesses all the information about exposure, test results, and contact information [42]. As a result, centralized systems are more suited to provide a "bigger picture" of the movement and spread of the virus and transmission data than decentralized systems [15].

Accuracy: Contact tracing technologies are perceived as more accurate then manual tracing alone [32]. However, most of the digital contact tracing technologies are also innately error prone. In the context of proximity-determining applications, there are two types of errors that can occur: 1) false positives, or when the app falsely identifies a user as exposed when they were not, and 1) false-negatives, or when the app fails to detect an exposure to a disease [26]. Saxena et al. estimated error rates between 7-15%, including both false positive and false negative errors when using the Bluetooth "handshake" method [47]. Despite this accuracy issue, digital contact tracing apps notify the identified close contacts without human verification process. In this work, we investigate how the imperfect contact information identified by digital systems can be effectively verified or disputed during a manual contact tracing phone call. Below, we describe more background work around this topic.

2.3 Decision Making with Imperfect Data

While diverse types of data and technologies have been suggested to aid contract tracing efforts, researchers have discerned that data and technologies are ultimately imperfect for identifying close contacts [49]. For example, Park et al. investigated proximity studies in HCI and Ubicomp and pointed out that, in some contexts, the locations of smartphones may not be a good representation of the users' locations [38]. Such imperfections can be understood with the concept of the socio-technical gap, which refers to the gap between "social requirements and technical feasibility [10]." In other words, there is a gap between the previous proximity studies and the CDC defined concept of close contact; the existing proximity studies have explored computational ways to identify the proximity of two different devices, however, identifying the CDC defined concept of close contact poses other challenges such as the layout of the environments (e.g., walls, stairs, and other physical or logistical barriers) and the context of the social encounters (e.g., studying in the library, shopping in a grocery store, or eating in the cafeteria).

Previous HCI researchers have investigated how people take advantage of data when there is an inevitable socio-technical gap, especially in the labor movement context [29], human rights activism [12], resident-led initiatives to improve neighborhoods [34], and homeless services [27]. Notably, Garcia et al. investigated how human right activists in Mexico collaborate to utilize absent and conflicting data to address social safety issues [12]. We echo Garcia et al. in that the alliance between stakeholders is the most important tenant in overcoming the imperfection of data in real world settings. Based on the lessons learned in the previous data practice studies, in this study, we explore technologies that can help stakeholders (voluntary contact tracers and community members) to utilize one type of imperfect but promising data to address a public health issue.

Related to data practice studies, recent scholarship in human and Artificial Intelligence (AI) collaboration investigated how people can navigate the imperfect results of an AI system when they make decisions [13]. Although our contact tracing visualization tool is not using AI, the process of contact tracers and community members making sense of imperfect WiFi data on-the-fly to make decisions about who should be counted as close contacts is similar to the way human interacts with limitations in AI algorithms for decision making. Specifically, Cai et al. elaborated on imperfect algorithms in the medical context describing that "no algorithm can perfectly capture an expert's ideal notion of similarity for every case [17]." Therefore, users of AI systems developed their own strategies to "disambiguate" data, algorithms, and the results of AI systems. These human-centered approaches to human-AI collaboration or human-data collaboration shed light on the importance of contextualized AI systems or data in the existing work practices [18]. Inspired by those human-centered AI studies, we investigate how contact tracers can make decisions using the WiFi proximity data of community members. We focused on understanding the collaboration between volunteer contact tracers and WiFi data as well as between volunteer contact tracers and community members to contextualize our proposed technology in existing contact tracing practices.

By aligning data practice studies and human-AI collaboration studies, we set our design goal as investigating the potentials of centralized forms of WiFi data in addressing public health concerns. We acknowledge that WiFi data can be an imperfect aid to contact tracing efforts and can bring privacy tensions, however, our study can contribute to the understanding of the collaboration processes between contact tracers and imperfect data in emergency situations and future designs of technologies to aid contact tracing efforts to contain pandemic situations.

3 USER-CENTERED DESIGN OF DIGITAL CONTACT TRACING VISUALIZATION

We developed a digital contact tracing visualization with the goal of capturing and displaying University community members' locations over selected periods of time. We studied a university campus because it provides both a readily accessible WiFi infrastructure and an established contact tracing process with trained contact tracers. The University that was studied is in an urban setting and the majority of the community members utilize the WiFi network for large parts of the day. In this section, we first explain the model used to collect community members' WiFi data, and then the design process of the visualization tool. All methods and procedures explained in the following section were approved by the Institutional Review Board (IRB) prior to conducting the study.

3.1 WiFi Location Data

Our study collects data using centralized proximity determination via the University's campus WiFi network. WiFi was chosen as the source of data for this study due to its ability to provide locations and duration of co-located University members. Furthermore, in contrast to approaches that require users to install an application on their device, the infrastructure-based approach reduce the user's burden as they likely use the campus WiFi network in their everyday lives.

When University community members have their wireless devices connected to the campus WiFi network, these devices are automatically authenticated with the network. Logs of these authentications are kept for various maintenance and troubleshooting purposes, and include access point identifiers, device identifiers, and date and time stamps. In order to determine proximity with this data, we examined the logs and determined which access point a user's device is connected to. The physical location of the access point to which a device is authenticating is a proxy for the approximate location person who is using that device. In addition, the amount of time a device stays continuously authenticated at a single access point can be inferred as the amount of time they spent at that location.

In the approximate 250 buildings on this campus, there are over 7000 WiFi access points. The high density of access points on the campus network make this method of determining co-location and proximity reasonably accurate. This method is effective at providing location and duration of co-locations, and, because most University community members use the University WiFi, the adoption rate is high (>90%). For contact tracing purposes, these are desirable characteristics as proximity and duration of contact are essential to determining close contacts.

One limitation associated with using centralized proximity determination in this way is the potential to report false-positives. Because WiFi access points can cover large areas, there is potential for two co-locators to not actually come into close contact with each other. In addition, false-negatives are possible because if a community member has no device connected to WiFi or is out of range of the access points, their data cannot be captured.

As was noted in the background section, proximity determining systems where the data is stored in a central database can raise privacy concerns among community members. In addition, infrastructure based proximity determination is location data, and people are less comfortable with the idea of their locations being made public than who they have been in contact with [31]. Despite these concerns, we believe the benefits of this approach such as ease of access and infrastructure availability still offer a promising opportunity to apply this to manual contact tracing.

3.2 Design Process

In order to design a technology to support University contact tracers, we employed a human-centered design approach. The team initially conducted interviews with contact tracers and developed personas and storyboards to inform the problem space and determine where the current contact tracing process falls short in identifying close contacts of positive cases (Figure 1-a). Following this, we conducted evaluation sessions and used the feedback to refine the tool's design (Figure 1-b). Below we explain the rationale behind the final visualization, and how it was uniquely designed to support contact tracers and help them achieve their goals.

3.2.1 User Research & Initial Design. To understand how University health systems approach contact tracing, and what opportunities and challenges might arise for designing a digital tool in the manual contact tracing process, we first interviewed 4 health authorities from two different universities with established contact tracing procedures. From one university we interviewed the Director of Health Services, the Clinical Lead for Contact Tracing, and the Contact Tracing Supervisor, and from the second university we interviewed the Director of Health Services. These authorities were in charge of recruiting contact tracers, setting up training, and implementing guidelines and processes respective to their Universities. The interview questions were focused on understanding how a university contact tracing system functions, and what opportunity there is for the application of a new technology to support the University's effort. From the preliminary interviews, we found that the university health authorities were encouraged by the idea to alleviate challenges of contact tracing, but concerned with establishing trust between the community members and a tool that could potentially have identifying information, and the tool's ability to capture off-campus contacts.

With these higher level focuses in mind, we conducted further interviews with two trained, active University contact tracers in order to understand their workflow when conducting a contact tracing call. We found that a common challenge that contact tracers face in their current process included how to identify close contacts that the positive case has a difficult time recalling. This concept is depicted in a story board (Figure 1-a) that demonstrates the issue in the context of contact tracing, and served as a starting point for a digital solution to the problem.

To address the challenge of community member's recall, the initial design of the visualization focused on highlighting the information of co-locators based on locations and duration that they shared with a positive case (Figure 1-b). For example, data from the WiFi authentication logs was visualized as a table of colored blocks, with time on the X-axis, much like a Gantt chart [19], and co-locator information on the Y-axis. Location was encoded as color, and duration was encoded as the length of the block. The top row of the table was the positive case, and subsequent rows of the visualization were co-locators over a period of time, determined by a date filter. If a person had co-located with the positive case, a block would indicate for how long and in what type of location directly below the corresponding block of the positive case.

An important part of contact tracing is understanding the infectious period of the positive case, as it determines the timeline of interest for contact tracers. This study was conducted between May and August of 2021. At the time of the study, according to the CDC, the infectious period, or period of time when the positive case is capable of infecting others, is defined as starting 2 days before someone becomes symptomatic or, for asymptomatic persons, 2 days prior to positive specimen collection [5]. If able to determine this information, contact tracers can focus their conversations to specific days and make a list of close contacts during those days in the case investigation. Therefore, our initial design has the data visualized on a timeline, and the intention was to give contact tracers enough data to help community members that have forgotten co-location events or close contacts.

3.2.2 Initial User Feedback & Final Design. The initial visualization design was demonstrated via high-fidelity prototypes to the 4 university health authorities and two contact tracers that were previously interviewed for the purpose of feedback on the design. The participants acknowledged that the information presented on the tool such as time-stamped location data and potential co-locators has the potential to jog student memory. In addition, the contact tracers in particular felt that knowing the general number of interactions that a case has is beneficial before a call for preparation purposes. However, participants were concerned that the visualization could present a privacy concern for people within the community. For example, in the initial design, identifying co-locator information, such as name and University ID, is visible in the tool. In addition, the contact tracers claimed that there were specific locations of interest that they considered at higher risk of close contact, or "high case potential" (i.e., an indoor cafe is higher risk than a outdoor seating area), and that they would like to group by those areas.

Based on the feedback from the participants that were shown the initial tool design, we addressed the privacy concern by changing the tool design to show de-identified information of people who tested positive and their co-locators. Both the positive cases' and potential co-locators' name and University ID in the tool were replaced with masked case ID's that cannot be tracked back to either community member. Only the contact tracer, who enters the positives case's masked ID at the top of the tool, knows the identity of the positive case (Figure 2-1). Because the de-identified version of the tool can still serve the purpose of supporting the contact tracing process, for example, through memory jogging, we intend for the de-identified design to be used in practice by contact tracers as well as in the study. Additionally, to address the request for "high case potential" areas, interaction functionalities were added such as location category filtering and co-location sorting (Figure 2-2). To sort the data, the contact tracers could click on the positive case's data of interest in the top row, and the visualization would update

(a) Story Board

(b) Initial Prototype (All names in this figure are mock names)

Figure 1: (a) The storyboard depicts a community member having trouble recalling his close contacts, but the contact tracer uses the visualization tool to help jog his memory. (b) The initial prototype design shows time and co-locators on the X- and Y-axes, respectively, and duration of co-location visualized as colored blocks. Co-location details and categorization are shown in the right panel.

to show all co-locators in that location at the top of the table in descending co-location duration.

The above mentioned design changes were presented to the same group of participants, and the feedback was positive. The contact tracers stated the need for flexibility in the timeline shown, because on a call, the infectious period information can change. The final tool design (Figure 2), included date input functionality (Figure 2-1) to allow the contact tracer to edit dates shown within the view. Finally, the meta data associated with a co-location event such as building type and duration is shown in the panel above the table to further contextualize the data visually for contact tracers (Figure 2-3). This design was carried over into the final visualization tool (Figure 2-4).

4 EXPLORING THE VISUALIZATION TOOL IN PRACTICE

In order to explore the visualization tool in use and understand the implications of using WiFi location data as a means of supporting contact tracing efforts in a University setting, we conducted mock contact tracing calls with contact tracers and community volunteers. We received IRB approval for this user study. The mock calls were conducted by providing previously trained contact tracers with the visualization tool, and community volunteers acting as "positive cases" intended for the tracers to investigate. All data in the visualization tool was de-identified using masked case ID's for both the "positive case" and the co-locators in order to protect privacy. We then conducted interviews with both contact tracers and community members to surface participants' attitudes, opinions, and perceptions of the experience. The mock contact tracing calls took an average of 13 minutes; the follow-up interviews with contact tracers and community members took about an average of 25 minutes. We compensated a $25 Amazon gift card to the contact tracer participants and $15 Amazon gift card to the community member participants. The contact tracers were compensated more than the community members because they conducted multiple calls and therefore volunteered considerably more time during the study than the volunteer community members, who only participated in one call each.

4.1 Participants

To recruit community member participants for the user study, an email survey was sent out to every department mailing list of the University explaining the purpose, procedure, and privacy conditions of the study. The email stated that consent to participate did not give the researchers access to browser history or information, but only to logs of which WiFi access points that participants' devices are connected to. In addition, email recipients were assured that their data would be de-identified and only used for this study, and that the research team will only be able to identify individuals if they consent to participate. Willing volunteers that responded to the email were then contacted and scheduled for study sessions. To recruit contact tracers, we contacted the clinical lead of the University contact tracing program. She shared our recruitment details to all of the contact tracers who were volunteering for contact tracing at the University. In total, 5 contact tracers (4 female) and 14 community volunteers (7 female) were recruited to participate. The breakdown of each group is shown in *Table 1: Participant Demographics*. Of the 14 volunteers, 9 were students (6 female) that resided on-campus, 4 were staff members (one female) that resided off-campus, and one was a male faculty member that also resided off-campus. Note that none of the participants from the previous design study were recruited to participate in the mock contact tracing calls. Data was analyzed immediately after every mock call, and participant recruiting ended when recurring themes were recorded in the data.

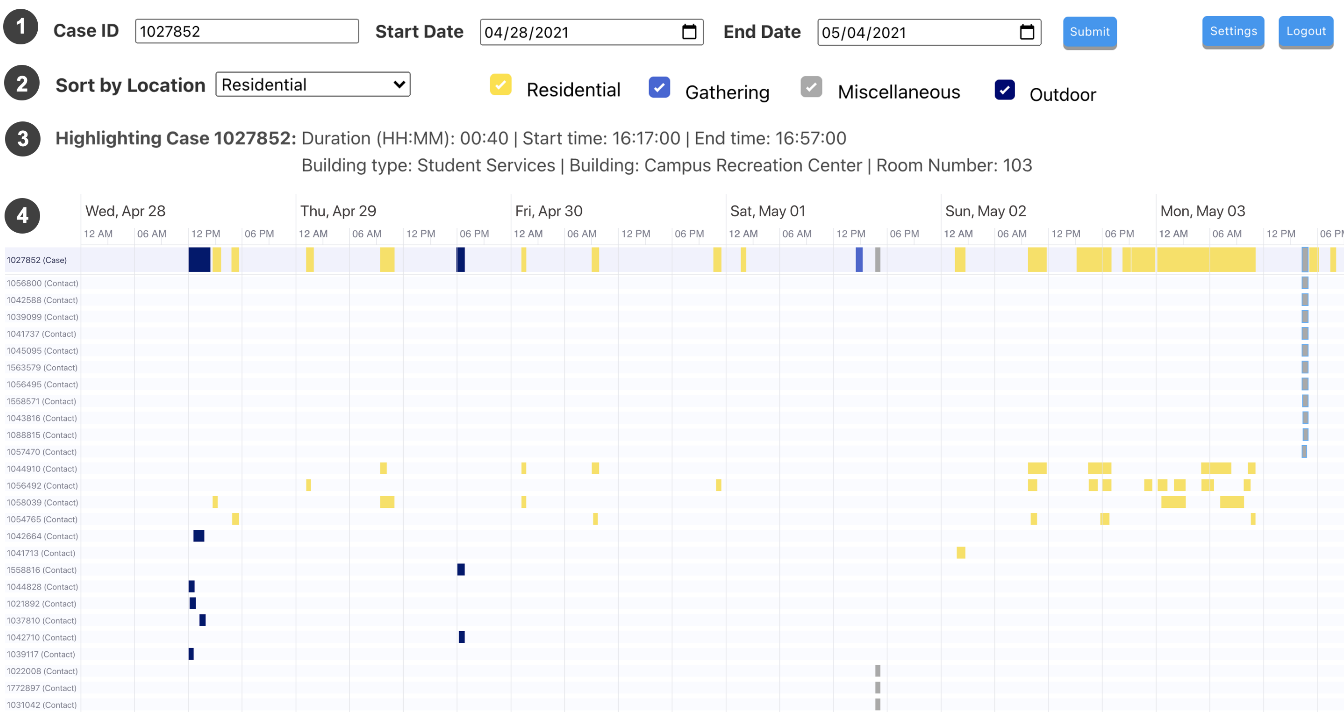

Figure 2: The final design of the visualization tool includes (1) the input fields for a community member's case ID and their symptom timeline, (2) sorting and filtering functionalities, (3) information associated with a co-location event, and (4) WiFi data visualization

Table 1: Participant Demographics

Group	Code	Gender	Description
5 University Contact Tracers	CT01 - 04	Female	University contact tracing volunteer
	CT05	Male	University contact tracing volunteer
14 University Community Members	CM01, CM04, CM05, CM07, CM09, CM10	Female	University student, resides on-campus
	CM02, CM03, CM18	Male	University student, resides on-campus
	CM06, CM11, CM14	Male	University staff, resides off-campus
	CM13	Female	University staff, resides off-campus
	CM12	Male	University faculty, resides off-campus

4.2 Preparation for mock-up contact tracing phone calls

Prior to conducting the contact tracing calls, de-identified case ID numbers were created for each volunteer community member and any co-locators shown the tool. The contact tracers were given the de-identified numbers for testing purposes and no identifiable information was shared between the tracer and volunteer regarding any close contacts identified during the call. In order to comply with privacy restrictions of the study, login access to the tool was not given to the contact tracers. Instead, tracers were able to access the tool from the researcher's desktop via the screen control function of the Bluejeans meeting platform.

Each contact tracer was provided with a short video that demonstrated the core functionalities of the tool intended for the contact tracing procedure, and then walked through how to request screen control from the researcher. To help contact tracers become familiar with the tool, the researcher asked them to complete the following five tasks with mock "positive case" data: 1) load mock student data into the tool and set it for a specific time period, 2) use the tool to identify co-locators to the "positive case" during a specific day and time, 3) sort the data to show co-locators that spent the most time with the "positive case", 4) sort the data to show co-locators from residential areas, and 5) reset the data in the tool. After these tasks were completed, tracers were allowed to explore any additional functionalities of the tool, ask questions, and load the community volunteers' data prior to the start of the call.

4.3 Mock-up contact tracing phone calls

Contact tracers were asked to participate in at least 2 mock contact tracing calls. Due to difficulty recruiting available contact tracers, one contact tracer volunteered to conduct 4 calls. Each community volunteer participated in one mock call, resulting in 14 total contact tracing calls. Prior to each call, community volunteers were given fabricated dates that they started showing symptoms, received a COVID19 test, and were notified of a positive test result. The volunteers would use this hypothetical data to act as a "positive case," and interacted with the tracer as such. Note that contact tracers utilized the tool and its data while conducting their normal contact tracing process with each "positive case." Although the COVID-19 test result is fabricated, we used real location data of community members and they engaged in conversation with the contact tracer based on their actual movement history and close contacts.

Community volunteers were given specific times to call into the meeting via phone, and were not allowed to view the tool being used during the contact tracing call. When the community volunteer called in, the contact tracing call would begin. The structure of the call was intended to simulate a real-life contact tracing situation, where the tracers used the same script and process as if the community volunteer had actually tested positive. Each call was between one contact tracer and one community volunteer.

4.4 Follow-up interviews

Immediately following the mock contact tracing calls, we conducted semi-structured interviews with contact tracers. These interviews were intended to assess the tracers attitudes and perceptions of the tools usability, the challenges and benefits it presents to the contact tracing process, and data privacy implications. To investigate usability of the tool, questions focused on the process that the contact tracers usually use during their calls, and how the tools functionalities aid or hinder that process. When assessing challenges and benefits of the tool, contact tracers were asked what aspects of the tool and the data presented were challenging, and if the tool helped identify additional close contacts. The data privacy-focused questions were mainly focused on contact tracers' perception and comfort level with asking the "positive case" about data that they can see in the tool.

Separate semi-structured interviews were conducted with each community volunteer after they participated in their mock contact tracing call. The interviews were focused on understanding the experience of contact tracing with their data visible to the tracer, their perceptions of data privacy, and their thoughts on public safety. During the interviews, the community member was shown the tool visualizing their own WiFi data. With the tool, the community members were asked questions about their thoughts on their data being used in this way for contact tracing, and if the visualization accurately displayed their location data.

4.5 Data analysis

All mock calls and interview sessions were video-recorded. The recordings were transcribed and analyzed for qualitative and empirical data.

In order to answer our research goal of understanding benefits and challenges that the data visualization introduces to the contact tracing process, we used thematic analysis [16, 24] and open coding to evaluate and categorize inferences from research participants. The transcriptions of both contact tracers and community members' interviews were reviewed line by line by the first author, and any references to attitudes, feelings, or thoughts about the tool were coded. Following this, two other authors reviewed the codes with the first author and then grouped the codes together based on common themes until a consensus was reached between all three authors on the data. Codes were grouped based on the participant's attitude in relation to the tool. For example, if the context of the code was something negative about the tool's impact on the contact tracing process or the tool itself such as inconsistencies in the data streams, the code would be grouped with similar inferences, the groups would be labeled, and associated as a challenge. Conversely, if the participant inference was something they liked about the tool or thought was helpful to the process such as the ability to facilitate memory jogging, the grouping process would be repeated, and the groups would be associated as a benefit.

The thematic analysis process resulted in three study themes that framed the rest of our analysis and discussion. The theme for benefit was determined to be promoting contact tracers confidence, and the themes for challenges were privacy and inconsistencies. Groups of sub-codes were organized under these themes. For example, contact tracers and community members felt that the data visualization would benefit the process by promoting context and memory jogging, and the contact tracers in particular felt more confident in their identification with the added context from the tool. These codes were grouped under promoting contact tracers confidence. Conversely, WiFi data ambiguity was often reported as being challenging for participants to reason, therefore inferences such as this were grouped under inconsistencies. Most of the codes that were classified into the groups and categories were commonly found by both groups of participants. However, one category—promoting contact tracer's confidence—was only mentioned by the contact tracer participants.

The mock calls were analyzed for empirical data by reviewing each call and logging specific data of interest as determined by the qualitative data and research questions. Data of interest included call lengths, number of discussions of specific event data, number of contacts identified on each call, and number of contacts identified in the tool and not in the tool.

5 RESULTS

Over the 14 mock phone calls conducted during the study, there were 75 total instances where volunteer community members and contact tracers discussed specific events, locations, and/or close contacts, and 158 total close contacts were identified, resulting in approximately 11 contacts identified per call (SD=11, min=1, max=43). There was an average of 5 close contacts validated in the tool and average of 6 close contacts validated outside the tool. Average call time was 13 minutes with the longest being 30 minutes and the shortest being 6 minutes (SD=7).

In this section, we first describe behaviors of interactions with the data that the visualization enabled. We then report 3 types of inconsistencies between the visualization and community member accounts: 1) WiFi data ambiguity, 2) community member's memory, and 3) information sharing hesitancy. For each inconsistency, we present how contact tracers and community members resolved the inconsistency, and the risk of leaving them unresolved. Finally, we report accompanying findings on how study participants perceive privacy and public safety.

During the mock contact tracing calls, contact tracers could interact with the data by either validating co-location instances or disputing them. Validation occurred when the community member confirmed that a co-locator displayed in the visualization was a close contact. Disputing data occurred when the data in the tool, either the location data or co-locator data, did not match with the community members account. When data was disputed, it was referred to as an inconsistency. We refer contact tracer participants as CT and community members as CM.

5.1 Inconsistencies

One consequence of using WiFi data in the contact tracing process is the potential for data mismatches, or inconsistencies. An inconsistency occurred when a community member disputed the data in the visualization. Inconsistencies occurred for three reasons: 1) the community members account did not match the data in the visualization due to WiFi data ambiguity, 2) the community member did not remember a particular co-location event displayed in the visualization, or 3) the community member did not want to share information with the contact tracer.

Of the 75 instances where a contact tracer and a community member discussed an event, 54 (72%) were associated with an inconsistency. The remaining 21 (28%) discussions of events were not associated with an inconsistency, and the data in the visualization matched the community member's recalled account exactly. Of the 54 inconsistencies, 52 (96%) were resolved between the contact tracer and the community member. To address the inconsistencies, participants on the call would discuss specific data about the co-location event and come to a logical conclusion about the nature of the inconsistency. For example, in a cluster-event including false positives, the community member and contact tracer would reason why the cluster occurred, and record the correct number of close contacts. Inconsistencies were typically resolved by either the community member correcting the data in the visualization, or the contact tracer using data in the tool to jog the community members memory. An inconsistency remained unresolved when, after discussion, the information could be neither validated or disputed. Below, we report three different types of inconsistencies that surfaced during the mock calls and how they were addressed by the call participants.

5.1.1 WiFi Location Data Ambiguity. One source of inconsistency was due to the centralized infrastructure-based proximity determination model. Because the data is captured via WiFi access points, anyone with a device connected to the same access point as the positive case will be displayed in the visualization as a co-locator. WiFi access points cover large areas, and multiple people that are not in close contact as defined by the CDC can be connected to the same one.

The result of this was that during the mock calls, many community members reported fewer actual close contacts than the visualization displayed. The remaining co-locators in the visualization that were not validated were false-positives. False positives, in the context of contact tracing, are when someone is reported as exposed when they were not. Of the 54 inconsistencies during the study, 24 (44%) were due to the tool reporting false-positives.

This type of inconsistency was usually resolved by the community member correcting the data by confirming their location and reporting the actual number of close contacts. For example, during one call, a contact tracer asked a community member about an instance in the visualization that showed the community member in the library (such as Figure 2). The visualization displayed more than 10 co-locators, but the community member only validated one as a close contact. Upon addressing the inconsistency during the call, the participants came to the conclusion that because it was a large, open study space, there could have been several people connected to the same WiFi access point. In this instance, the tool displayed many false-positives, but the inconsistency was resolved. False-positive cluster-events, or events where high density of people connected to one WiFi access point, occurred in communal buildings such as campus recreation centers, libraries, and study halls.

Converse to reporting false-positives, the visualization also introduced instances of false-negatives, or close contacts identified by the community member that were not captured in the data visualization. We found two reasons for false negative instances in our study. The first reason for false negatives was if a community member, either a positive case or co-locator, did not have their mobile device connected to the campus WiFi. The second reason was when community members had co-location instances off campus, out of range of the campus WiFi. In both cases, because the community members' mobile devices were not authenticating with the network of access points used in this study, their location data could not be captured. Of the total 75 instances of close contacts discussed between the contact tracer and community member, 18 (30%) were cases where the community member recalled an event that was not captured in the visualization due to being off campus or not having their WiFi connected.

There was a evidence that the occurrence of these inconsistencies reduced contact tracer trust in the visualization. For example, one contact tracer felt that he *"can't rely 100% on the tool [CT02]"* because the data is not always accurate, therefore more context is needed to identify close contacts. Another contact tracer claimed that because the tool is limited in capturing a very granular level of proximity, she would not rely solely on the tool for identification of close contacts, but use it to verify or complement the human's account [CT03]. These findings show that contact tracers are aware of the data ambiguity within the tool, and it impacted their trust in using the visualization to identify close contacts.

5.1.2 Gap in Community Member's Memory. The second type of inconsistency occurred when the community member forgot or mis-remembered an event. This happened in two ways: 1) the community member was asked for a recalled account of their movements and contacts and they forget an event in the report, or 2) a contact tracer prompted the community member with information and the community member initially did not recall the event.

To resolve this type of inconsistency, contact tracers most often tried to jog the community member's memory by prompting them with details from the visualization. During one mock call, the contact tracer used the sort function to better visualize a large co-location event in the community members data (see Figure 2 on Monday, May 3rd, approx 3pm). The community member at first claimed that he had stayed at his dorm that day, so did not know what the data could be. When the contact tracer gave the student a specific time, building name (Campus Recreation Center), and duration of the event, the student remembered going to the gym. In this case, having specific building data helped to jog the community member's memory, and they were able to come to a logical conclusion in determining the large number of co-locators were false-positives.

Both contact tracers and community members reported that misremembering events could happen often, as it is difficult to recall every interaction one has over several days. Of the 54 inconsistencies, there were 10 that were resolved by the contact tracer jogging the community members memory. In all 10 instances, the contact tracer was able to prompt the community member with specific dates, times, building names, or number of co-locators to assist their recall. Contact tracers and community members perceived that the contact tracers ability to jog memory with specific data as a beneficial aspect of the visualization. One community member recalled that they *"[...] tested positive last fall and it was very difficult to try to remember who I was around. [...] I guess the benefit of being able to more easily remember what I was doing since I seem to have forgetfulness of that at certain points [CM02]."*

The effective use of the visualization tool to prompt and help jog community members' memory to resolve memory gap inconsistencies is evidence that the visualization of WiFi location data is useful for manual contact tracing. In situations where the community member may not remember all of their interactions during their infectious period, the visualization showed promise as a way to fill the memory gap, thus addressing the well-known recall problem.

5.1.3 Information Sharing Hesitancy. The third cause of inconsistency, though not encountered during our mock contact tracing calls, is the reluctance of community members to share close contact information. Although all community members in this study were willing to share their close contact information and reported they would do so for a real contact tracing scenario, many of the contact tracers referenced previous experiences where a community member who had tested positive did not want to share the names and information of the people they had been in close contact with. Notably, one reason for this is due to feeling guilty about disrupting the life of their friends or colleagues to quarantine. The community members had similar reports. One community member reported that *"I had to go through this whole [contact tracing] thing last fall and I just felt bad at certain points to [have to] tell people I was with. And then their whole lives [were] interrupted [CM02]."*

The result of this is no data is addressed and no co-locators are validated as close contacts. As with other unresolved inconsistencies, contact tracers claimed remaining neutral and non-accusatory was important if faced with this, or any, type of inconsistency. For example, a contact tracer said *"You're trying to keep a good rapport with [community members]. You're trying to be his friend and his ally, and you want them to trust you. So if you start saying, 'well, I don't think you're telling me the truth here.' or 'I don't think you're right.' Then, it might jeopardize that [CT5]."*

5.1.4 Unresolved Inconsistencies. Of the 54 inconsistencies that arose during the study, two remained unresolved. In both instances, the inconsistency could not be resolved because the community member claimed to not be in the location that the visualization showed at the time that it showed. After discussing the data, the community member and contact tracer could not come to a logical conclusion as to what caused the inconsistency.

When inconsistencies arise, contact tracers felt they had to trust either the community member account or the data in the visualization. Some contact tracers were more inclined to trust the community member, while others were more inclined to trust the data. When asked what data to trust when addressing inconsistencies, one tracer claimed *"Assuming that they are telling the truth and that they might be forgetting something and if they're not maybe it's an error with the computer... just trusting the student [CT1],"* and another said *"You can't rely 100% on the tool, it just aids the case and their memory [CT02]."* Conversely, other contact tracers perceived the specific details in the tool to be more reliable: *"[Without this tool] we would just have their word and their information, we would not have actual facts [CT02]."* However, instances where co-locators were validated most "trusted" data by the contact tracers [CT4].

5.2 Privacy

The addition of WiFi location data to the contact tracing process elicits concerns regarding data privacy, as many people are *"very private about their activity and presence on the internet [CM05]."* For example, one community member said *"there's a lot of debate about location services and how moral it is for people to have access to that"* and that she is *"paranoid about [her] data being used for the wrong things [CM07]."*

Many community members claimed that because the study was clear in informing them of what data was going to be used and that the data would be de-identified, they were more willing to participate than if that information had not been provided. For example, CM01 said, *"the main thing is just to make sure people are OK with it and they know exactly what it's being used for [CM01]."* These findings suggest that the explicit communication of data that will be collected, and demonstrating how it is presented in the visualization, alleviated their privacy concerns associated with sharing their data for use in the visualization tool. For example, after being shown the tool, one student claimed that she would consent to providing her data because *"they're not going to [...] have all the information, like who you've been texting or what you've been searching [CM05]."* The transparency offered in this study enhanced the community members trust in how their data will be used.

Some community members expressed that they feel they should have access to the visualization themselves. The reasons for this varied by community member. One community member wanted access to the tool for the purpose of data agency and transparency. She felt that if she wanted to revoke permission to share her data, she could check the tool. *"...If I reached a point and felt like I had* *changed my mind [...] I could log in and see it's blank. I think that transparency would be really valuable [CM07]."* Additionally, other community members felt that because it is their data, they have a right to view it, claiming that *"people should be given free access to their own data [CM10]."*

5.3 Promoting Confidence of Contact Tracers

The contact tracers felt that being able to see location data and number of co-locators before the call in the visualization was particularly useful because they could prepare their approach for the call, estimate call time, and evaluate "potential level of risk [CT04]." In a regular contact tracing call, the tracer may only have the positive cases' name or ID number. *"Being a contact tracer, you are given a name and phone number. [...] you don't know what you're stepping into [CT02]."*

When the tracers were allowed to view the community members' data using the visualization before the call, they often identified instances of interest that they noted as important to ask. This was helpful for tracers to estimate a number of close contacts and aid in *"management and distribution of cases [CT02]"*. One tracer mentioned that if a community member had a particularly high number of co-locators, the contact tracing team might split the follow up calls between them to better manage their time. *"It could be a ten minute call or it could be a ten hour call. Especially when people lived in fraternity and sorority houses [CT02]."* In this case, when a student has many co-locators or is deemed "high risk" based on their movements or living situation, viewing the data before the call can allow the contact tracer to be better informed and more deliberate, and potentially work through a lot of data in a short period of time.

Moreover, the visualization helped to give more context to contact tracers. In one case, the contact tracer stated that she was not familiar with campus, so she would sometimes have difficulty when conversing about locations with community members. *"It's nice that you can see where [people] were. I'm not familiar with the campus, but for me to be able to pin point [a location] for them a little bit more helps [CT03]."* Having access to categorized location data during the mock calls allowed this tracer to be more confident and have a more productive conversation with the community member. For example, being shown the name and type of building will allow the contact tracer to prompt with more specific information, and understand the nature of events that are being discussed.

6 DISCUSSION

In this section, we present our findings in the context of socio-technical gaps in sensed-data systems, and discuss how these findings can inform the design of technologies that are developed to support decision making by making sense of imperfect data. Further, we discuss strategies that organizations should consider to address privacy and trust issues that can arise when deploying such systems in practice.

6.1 Socio-Technical Gaps

By providing the WiFi location data visualization to study participants as a means for working together to achieve the goal, we shifted the manual close contact identification process to computer supported cooperative work (CSCW) [37]. In doing so, we expose and investigate the socio-technical gaps [10] that arise as a result of the CSCW-adaptation to manual contact tracing.

During the mock contact tracing calls, we observed examples of socio-technical gaps. Most commonly, the gaps were due to imperfect data within the system. For example, the occurrence of false positive and false negative reports were gaps that resulted from the spatial resolution of the data. Due to the access points in the University WiFi infrastructure covering large areas, false-positive reports were common in large buildings such as the library or gym. In these large buildings, many people who are not in close contact can be connected to the same WiFi access point, thus showing as co-locators in the WiFi location data visualization. Alternatively, the high number of close contact identifications on the mock calls that were not captured in the data due to off-campus, out of range gatherings demonstrates that infrastructure based sensing is limited to the scope of the network itself.

In addition to the socio-technical gaps, we found community members' memory gaps and information sharing was also a challenge. Although these inconsistencies were present, contact tracers accomplished their goal of identifying close contacts of a community member by making sense of the the inconsistencies together with the community member. Below, we further discuss how the existence of socio-technical gaps in sensed data systems requires the users of the system to reason the data, or engage in sensemaking, in order to complete the decision making process.

6.2 Sensemaking of Imperfect Data

A significant body of HCI work in the health domain has explored strategies and tensions that arise when people make sense of imperfect data with others in collaborative settings [40, 41, 43, 54]. We contribute to the existing body of research by uncovering factors that facilitate and/or complicate making sense of imperfect data in the context of contact tracing. For example, our findings showed that visualizing the data that provides context such as building names and dates helped contact tracers jog community members' imperfect memory and confirm their account with the data. Therefore, it increased contact tracers' confidence in the subsequent close contact identifications. Moreover, the data showing location and number of co-locators boosted the confidence of contact tracers in assessing the relative risk and mobility of a positive case before the start of the call with a community member.

However, as was discussed above, there still exists the need to reason data that is unexpected, counterintuitive, or imperfect [41]. Before and during the mock contact tracing calls, the study participants engaged in collaborative sensemaking behaviors in order to understand and reason the data shown in the visualization, especially when more context was needed for an accurate close contact identification (e.g., in the false positive instance where a community member was only in close contact with one person, but the visualization showed many co-locators in a big building). Below, drawing from the collaborative sensemaking literature [40, 41], we discuss how study participants collaboratively reasoned the data that was provided in order to resolve gaps and confidently identify close contacts to the positive case. Further, we recommend design elements that can highlight and assist in collaborative sensemaking of imperfect data.

6.2.1 Designing for Sensemaking.

Due to the need to contextualize, reason, and make decisions with imperfect data, visualization systems should be designed specifically to support sensemaking of data. Particularly in visualizations that are intended to be used for collaborative decision making, providing context to the user for interpretation is crucial [43]. In the case of our visualization tool, the inconsistencies and gaps need to be contextualized in order for the contact tracers and community members to confidently identify close contacts.

Prioritizing Relevant Data: One way of adding context to data that is common in collaborative sensemaking the use of prioritization [40]. We observed specific data being discussed by the study participants repeatedly, indicating that this data was a priority for identifying close contacts. The contact tracers tended to confirm locations via building name and category provided in the tool, and time of a co-location event. In addition, the number of co-locators presented in the data was frequently asked to the community member as a way to confirm the number of close contacts.

This observation is similar to previous work's finding on the characteristics of collaborative sensemaking that prioritizing pieces of data is key for decision making within a group. As people are given information to consider, they make judgements on its relevancy and therefore importance to the final decision [40]. Because the need to contextualize and make sense of imperfect data due to socio-techincal gaps is common in sensed-data technologies, these systems should be designed to make the relevant information more visual in the right context so that the collaborators can easily prioritize the data to resolve the inconsistencies.

Surfacing the Gaps: We also suggest adding design elements in technologies using imperfect data that detect and surface when data is at risk of inconsistencies or gaps. For example, in our study, areas such as the campus recreation center and library produced many false-positive situations due to the high density of people in the building. To address this, creating an indicator on the co-location instance in the visualization explaining that the instance is part of a cluster-event that could include false positives would be beneficial to reduce cognitive load and aid in quicker collaboration. Or, for a more interactive approach, a pop-up or modal that informs the user of potential risks. Highlighting this information is a way to alert the tracer that this data likely needs addressing, and supports the collaborative decision making process between the contact tracer and the community member by adding context to the potential inconsistency. Additionally, transparency within a system is crucial for building user trust in the data it is presenting [30]. By explicitly surfacing where the system is lacking, collaborators will have more confidence in using the system, as they will feel informed and prepared of the gaps they will likely face.

Refinement of Data: Consistent with previous work [15, 30], our result suggests the possibility that contact tracers may lose trust in a contact tracing visualization that produces false-positive or false-negative results regularly. Especially within systems where data is used for collaborative decision making, such as a contact tracer and a community member identifying close contacts, trust in the data is a key component of success [43]. "Data refinement" has been suggested as a technique to increase user's trust in the system by allowing users to make small incremental changes to results from a system in order to increase the utility of the results to the user [17]. By being able to actively interact with the system and refine and make sense of the results, users feel agency in being able to "guide" the system to return results that were more beneficial, and thus more trust in its output. Our findings of the conversations that lead to inconsistency resolution and decision of close contacts were aligned with the concept of "data refinement." Because the contact tracers and community members were able to address both recall data and visualization data incrementally, they could make small changes to the information being presented in the visualization in order to validate, dispute, or capture extra data. Therefore, we suggest systems that are designed to explicitly support this data refinement process by allowing contact tracers to mark, make notes, and interact with the data as they are actively on a call with a community member. For example, in the case of false positives, the contact tracer might benefit from marking all those co-location instances as such, or, in the case of false-negatives, contact tracers can refine the visualization by adding the co-location instances into the tool. This refinement concept can not only visually add context to the data, but it would also build more confidence in the system as it allows contact tracers agency over the data they are using.

6.3 Designing for the Community

For systems that utilize community member data for decision making, we found that trust needs to be established between the community and the authorities that use the data. In this section, we discuss ways to assuage the privacy concerns that people may have when sharing their data.

6.3.1 Communication & Privacy.

In addition to inconsistencies caused by the limitations of the technology, the visualization's data collection requirement can also be a potential driver for socio-technical gaps in the system. Infrastructure-based, centralized contact tracing systems decrease the individual's privacy and control of their data. In the context of passive data collection for health purposes, a review of the space reported that people are willing to share their sensed data within groups of people with similar health goals or with physicians, however, applications that require data collection see declining adherence specifically due to privacy concerns [48].

Because the uptake of contact tracing systems is crucial to their efficacy, if individuals opt out or do not consent to have their data collected, the tool becomes less useful [31]. For these reasons, when deploying systems that collect individual data, efforts should be made in order to make sure that the community is fully informed of the system and its use.

Especially for systems that collect user data, it is important to communicate to the community of intended users what data will be collected and how it will be used. We found that when people are aware of what their data is being used for, they can make more informed decisions about whether or not to participate, and they can assess for themselves if they are comfortable with sharing. In addition, opt-in or opt-out options should be designed into the system so community members feel comfortable with changing their mind and have agency to do so.

6.3.2 Community Access to the Tool. There are many contact tracing systems such as NOVID and TraceTogether built for community members by allowing them to collect data using their own devices [3, 21]. Therefore, to explore a new area of research–increasing contact tracing efficiency from an organizational context–we designed the visualization specifically for use by contact tracers by utilizing organizational infrastructure. Interestingly, our findings indicate that many community members felt that they would also benefit from having access to the visualization. For example, some community members felt that they should be given access to the data that they are providing, thus creating a sense of agency over the process and supporting transparency. Allowing community member access to the tool before a contact tracing call could also serve as memory jogging, because they will be able to see the data that the contact tracer would be using.

When considering whether or not to give the community members access to their data in the contact tracing technology, privacy of co-locators should be considered. Currently, contact tracers do not reveal to close contacts who the positive was, however, with co-location data there is potential for that information to be revealed. Additionally, it should be considered if community members have access to the tool at all times or only if they receive a positive test and require contact tracing.

7 LIMITATIONS

One limitation associated with this study is the small number of participants. Due to contact tracer volunteers needing specialized skills, the pool of volunteers was small. In addition, our study likely contains bias within the community members. In order for the community members to participate in the study, they had to consent to letting us use their location data in the visualization tool. Therefore, participants who were willing to participate in our study already felt comfortable with sharing their WiFi data for contact tracing purposes. This bias excludes community members that are not comfortable with sharing their data from participating in the study. Future studies should explore perspectives of those community members who are uncomfortable sharing their data for contact tracing purposes.

Lastly, the real-world application of this study is limited due to its utilization of the large, pre-existing network infrastructure available on an organization. Despite the limitations, our contribution of how to support collaborative sensemaking through the design of systems that use imperfect data still presents useful insights.

8 CONCLUSION

Our visualization tool designed for manual contact tracing successfully supported contact tracers in jogging memory of community members and boosting their confidence. However, we further found challenges in incorporating the visualization in the contact tracing process due to inconsistencies that arose between the community members' account and WiFi location data, as well as privacy concerns regarding using WiFi location data in contact tracing. In designing tools that make use of imperfect data, we recommend design elements that highlight data potentially at risk for inconsistencies, in order to help users prioritize and make sense of the data. We further suggest incorporating data refinement opportunities within the system in order to build trust between the user and the system and support confident decision making. Lastly, we present the need for organizations to be fully transparent in their collection and use of community member data, in order to better inform and potentially assuage any privacy concerns that the community may have when deploying a system that utilizes their data.

ACKNOWLEDGMENTS

We are grateful to the students and faculty involved with the CampusLife initiative at the Georgia Institute of Technology. This study was bolstered by the initial prototypes designed by Yiheng Qi, along with iterations from Nesha Prabahar, Samruddhi Kulkarni, and Maanit Madaan. The back-end of the tool was maintained thanks to Sonia Sargolzaei and Yexin Tian. The early collaboration and user research with contact tracers was thanks to the efforts of Shubhangi Gupta, Soumya Pachigolla, Tanuja Sawant, and Taylor Stillman. We also thank Duen Horng (Polo) Chau, Jay Wang, and Haekyu Park for their feedback on the visualization tool. Lastly, this project was only possible due to Matt Sanders, who facilitated our collaboration with the Office of Information Technology to ensure proper data acquisition, security, and maintenance. We also would like to acknowledge that Vedant Das Swain and Munmun De Choudhury were supported by NSF RAPID grant No. 2027689.

REFERENCES

[1] 2019. Appendices. https://www.cdc.gov/coronavirus/2019-ncov/php/contact-tracing/contact-tracing-plan/appendix.html#contact

[2] 2020. Singapore government launches new app for contact tracing to combat spread of COVID-19. https://www.mobihealthnews.com/news/asia/singapore-government-launches-new-app-contact-tracing-combat-spread-covid-19

[3] 2021. https://www.tracetogether.gov.sg/common/privacystatement/index.html

[4] 2021. Coursera contact tracing course. https://www.coursera.org/learn/covid-19-contact-tracing

[5] 2021. COVID-19 Contact Tracing. https://www.cdc.gov/coronavirus/2019-ncov/daily-life-coping/contact-tracing.html

[6] 2021. How TraceFi works. https://covidtech.harvard.edu/howitworks.html

[7] 2021. Operational Considerations for Adapting a Contact Tracing Program to Respond to the COVID-19 Pandemic in non-US Settings. https://www.cdc.gov/coronavirus/2019-ncov/global-covid-19/operational-considerations-contact-tracing.html#anchor_1624283937257

[8] 2021. proximity. https://dictionary.cambridge.org/us/dictionary/english/proximity

[9] MEDHA KALLEM | August 19. 2020. NOVID app makes contact tracing mobile. https://www.jhunewsletter.com/article/2020/08/novid-app-makes-contact-tracing-mobile

[10] Mark S Ackerman. 2000. The intellectual challenge of CSCW: The gap between social requirements and technical feasibility. *Human–Computer Interaction* 15, 2-3 (2000), 179–203.

[11] Nadeem Ahmed, Regio A Michelin, Wanli Xue, Sushmita Ruj, Robert Malaney, Salil S Kanhere, Aruna Seneviratne, Wen Hu, Helge Janicke, and Sanjay K Jha. 2020. A survey of COVID-19 contact tracing apps. *IEEE access* 8 (2020), 134577–134601.

[12] Adriana Alvarado Garcia, Alyson L Young, and Lynn Dombrowski. 2017. On making data actionable: How activists use imperfect data to foster social change for human rights violations in Mexico. *Proceedings of the ACM on Human-Computer Interaction* 1, CSCW (2017), 1–19.

[13] Saleema Amershi, Dan Weld, Mihaela Vorvoreanu, Adam Fourney, Besmira Nushi, Penny Collisson, Jina Suh, Shamsi Iqbal, Paul N Bennett, Kori Inkpen, et al. 2019. Guidelines for human-AI interaction. In *Proceedings of the 2019 chi conference on human factors in computing systems*. 1–13.

[14] James Benneyan, Christopher Gehrke, Iulian Ilies, and Nicole Nehls. 2021. Community and Campus COVID-19 Risk Uncertainty Under University Reopening Scenarios: Model-Based Analysis. *JMIR public health and surveillance* 7, 4 (2021), e24292.

[15] Isobel Braithwaite, Thomas Callender, Miriam Bullock, and Robert W Aldridge. 2020. Automated and partly automated contact tracing: a systematic review to inform the control of COVID-19. *The Lancet Digital Health* (2020).

[16] Virginia Braun and Victoria Clarke. 2012. Thematic analysis. (2012).

[17] Carrie J Cai, Emily Reif, Narayan Hegde, Jason Hipp, Been Kim, Daniel Smilkov, Martin Wattenberg, Fernanda Viegas, Greg S Corrado, Martin C Stumpe, et al. 2019. Human-centered tools for coping with imperfect algorithms during medical decision-making. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–14.

[18] Carrie J Cai, Samantha Winter, David Steiner, Lauren Wilcox, and Michael Terry. 2021. Onboarding Materials as Cross-functional Boundary Objects for Developing AI Assistants. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–7.

[19] Wallace Clark. 1922. *The Gantt chart: A working tool of management.* Ronald Press Company.

[20] Muawya Habib Sarnoub Eldaw, Mark Levene, and George Roussos. 2018. Presence analytics: making sense of human social presence within a learning environment. In *2018 IEEE/ACM 5th International Conference on Big Data Computing Applications and Technologies (BDCAT)*. IEEE, 174–183.

[21] People who care at Expii. 2021. https://www.novid.org/#how-it-works

[22] Christophe Fraser, Lucie Abeler-Dörner, Luca Ferretti, Michael Parker, Michelle Kendall, and David Bonsall. 2020. Digital contact tracing: comparing the capabilities of centralised and decentralised data architectures to effectively suppress the COVID-19 epidemic whilst maximising freedom of movement and maintaining privacy. *University of Oxford* (2020).

[23] Kara Gavin. 2020. Flattening the Curve for COVID-19: What Does It Mean and How Can You Help? https://healthblog.uofmhealth.org/wellness-prevention/flattening-curve-for-covid-19-what-does-it-mean-and-how-can-you-help

[24] Judith A Holton. 2007. The coding process and its challenges. *The Sage handbook of grounded theory* 3 (2007), 265–289.

[25] Jeremy Hsu. 2021. Contact Tracing Apps Struggle to Be Both Effective and Private. https://spectrum.ieee.org/contact-tracing-apps-struggle-to-be-both-effective-and-private

[26] Gabriel Kaptchuk, Daniel G Goldstein, Eszter Hargittai, Jake Hofman, and Elissa M Redmiles. 2020. How good is good enough for COVID19 apps? The influence of benefits, accuracy, and privacy on willingness to adopt. *arXiv preprint arXiv:2005.04343* (2020).

[27] Naveena Karusala, Jennifer Wilson, Phebe Vayanos, and Eric Rice. 2019. Street-Level Realities of Data Practices in Homeless Services Provision. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019), 1–23.

[28] Cristina Criddle & Leo Kelion. 2020. Coronavirus contact-tracing: World split between two types of app. https://www.bbc.com/news/technology-52355028

[29] Vera Khovanskaya, Phoebe Sengers, and Lynn Dombrowski. 2020. Bottom-Up Organizing with Tools from On High: Understanding the Data Practices of Labor Organizers. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–13.

[30] John D Lee and Katrina A See. 2004. Trust in automation: Designing for appropriate reliance. *Human factors* 46, 1 (2004), 50–80.

[31] Tianshi Li, Cori Faklaris, Jennifer King, Yuvraj Agarwal, Laura Dabbish, Jason I Hong, et al. 2020. Decentralized is not risk-free: Understanding public perceptions of privacy-utility trade-offs in COVID-19 contact-tracing apps. *arXiv preprint arXiv:2005.11957* (2020).

[32] Xi Lu, Tera L. Reynolds, Eunkyung Jo, Hwajung Hong, Xinru Page, Yunan Chen, and Daniel A. Epstein. 2021. Comparing Perspectives Around Human and Technology Support for Contact Tracing. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–15.

[33] Alison Mack, Eileen R Choffnes, Margaret A Hamburg, David A Relman, et al. 2009. *Microbial evolution and co-adaptation: a tribute to the life and scientific legacies of Joshua Lederberg: workshop summary.* National Academies Press.

[34] Amanda Meng, Carl DiSalvo, and Ellen Zegura. 2019. Collaborative data work towards a caring democracy. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019), 1–23.

[35] Australian Government Department of Health. 2020. COVIDSafe app. https://www.health.gov.au/resources/apps-and-tools/covidsafe-app

[36] World Health Organization et al. 2019. *Non-pharmaceutical public health measures for mitigating the risk and impact of epidemic and pandemic influenza: annex: report of systematic literature reviews.* Technical Report. World Health Organization.

[37] Thomas D Palmer and N Ann Fields. 1994. Computer supported cooperative work. *Computer* 27, 5 (1994), 15–17.

[38] Jung Wook Park, Hayley I Evans, Hue Watson, Gregory D Abowd, and Rosa I Arriaga. 2020. Growing Apart: How Smart Devices Impact the Proximity of Users to Their Smartphones. *IEEE Pervasive Computing* 19, 3 (2020), 79–88.

[39] Sangchul Park, Gina Jeehyun Choi, and Haksoo Ko. 2020. Information technology–based tracing strategy in response to COVID-19 in South Korea—privacy controversies. *Jama* 323, 21 (2020), 2129–2130.

[40] Sharoda A Paul and Madhu C Reddy. 2010. Understanding together: sensemaking in collaborative information seeking. In *Proceedings of the 2010 ACM conference on Computer supported cooperative work*. 321–330.

[41] Shriti Raj, Joyce M Lee, Ashley Garrity, and Mark W Newman. 2019. Clinical data in context: towards Sensemaking tools for interpreting personal health data. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 3, 1 (2019), 1–20.

[42] Elissa M Redmiles. 2020. User Concerns 8 Tradeoffs in Technology-facilitated COVID-19 Response. *Digital Government: Research and Practice* 2, 1 (2020), 1–12.

[43] Jessica Schroeder, Jane Hoffswell, Chia-Fang Chung, James Fogarty, Sean Munson, and Jasmine Zia. 2017. Supporting patient-provider collaboration to identify individual triggers using food and symptom journals. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*. 1726–1739.

[44] Ira B. Schwartz, James Kaufman, Kun Hu, and Simone Bianco. 2020. Predicting the impact of asymptomatic transmission, non-pharmaceutical intervention and testing on the spread of COVID19. (2020).

[45] Lucy Simko, Ryan Calo, Franziska Roesner, and Tadayoshi Kohno. 2020. COVID-19 contact tracing and privacy: studying opinion and preferences. *arXiv preprint arXiv:2005.06056* (2020).

[46] Vedant Das Swain, Jiajia Xie, Maanit Madan, Sonia Sargolzaei, James Cai, Munmun De Choudhury, Gregory D Abowd, Lauren N Steimle, and B Aditya Prakash. 2021. WiFi mobility models for COVID-19 enable less burdensome and more localized interventions for university campuses. *medRxiv* (2021).

[47] Yvonne Taunton. 2020. Smartphone-based automated contact tracing: Is it possible to balance privacy, accuracy and security.

[48] Alina Trifan, Maryse Oliveira, and José Luís Oliveira. 2019. Passive sensing of health outcomes through smartphones: systematic review of current solutions and possible limitations. *JMIR mHealth and uHealth* 7, 8 (2019), e12649.

[49] Amee Trivedi and Deepak Vasisht. 2020. Digital Contact Tracing: Technologies, Shortcomings, and the Path Forward. *SIGCOMM Comput. Commun. Rev.* 50, 4 (2020), 75—81. https://doi.org/10.1145/3431832.3431841

[50] Amee Trivedi, Camellia Zakaria, Rajesh Balan, Ann Becker, George Corey, and Prashant Shenoy. 2021. WiFiTrace: Network-based Contact Tracing for Infectious Diseases Using Passive WiFi Sensing. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 5, 1 (2021), 1–26.

[51] Viktor Von Wyl, Sebastian Bonhoeffer, Edouard Bugnion, Milo Alan Puhan, Marcel Salathé, Theresa Stadler, Carmela Troncoso, Effy Vayena, and Nicola Low. 2020. A research agenda for digital proximity tracing apps. *Swiss medical weekly* 150 (2020), w20324.

[52] Laurel Wamsley. 2021. Singapore Says COVID-19 Contact-Tracing Data Can Be Requested By Police. https://www.npr.org/sections/coronavirus-live-updates/2021/01/05/953604553/singapore-says-covid-19-contact-tracing-data-can-be-requested-by-police

[53] Shweta Ware, Chaoqun Yue, Reynaldo Morillo, Jin Lu, Chao Shang, Jayesh Kamath, Athanasios Bamis, Jinbo Bi, Alexander Russell, and Bing Wang. 2018. Large-scale automatic depression screening using meta-data from wifi infrastructure. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 4 (2018), 1–27.

[54] Peter West, Max Van Kleek, Richard Giordano, Mark J Weal, and Nigel Shadbolt. 2018. Common barriers to the use of patient-generated data across clinical settings. In *proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–13.

[55] Ye Xia and Gwendolyn Lee. 2020. How to return to normalcy: fast and comprehensive contact tracing of COVID-19 through proximity sensing using mobile devices. *arXiv preprint arXiv:2004.12576* (2020).