

4X: A Hybrid Approach for Scaffolding Data Collection and Interest in Low-Effort Participatory Sensing

KAPIL GARG, Northwestern University, USA
 YONGSUNG KIM, Northwestern University, USA
 DARREN GERGLE, Northwestern University, USA
 HAOQI ZHANG, Northwestern University, USA

Participatory sensing systems in which people actively participate in the data collection process must account for both the needs of data contributors and the data collection goals. Existing approaches tend to emphasize one or the other, with opportunistic and directed approaches making opposing tradeoffs between providing convenient opportunities for contributors and collecting high-fidelity data. This paper explores a new, hybrid approach, in which collected data—even if low-fidelity initially—can provide useful information to data contributors and inspire further contributions. We realize this approach with 4X, a multi-stage data collection framework that first collects data opportunistically by requesting contributions at specific locations along users’ routes and then uses collected data to direct users to locations of interest to make additional contributions that build data fidelity and coverage. To study the efficacy of 4X, we implemented 4X into LES, an application for collecting information about campus locations and events. Results from two field deployments ($N = 95$, $N = 18$) show that the 4X framework created 34% more opportunities for contributing data without increasing disruption, and yielded 49% more data by directing users to locations of interest. Our results demonstrate the value and potential of multi-stage, dynamic data collection processes that draw on multiple sources of motivation for data, and how they can be used to better meet data collection goals as data becomes available while avoiding unnecessary disruption.

CCS Concepts: • **Human-centered computing** → **Ubiquitous and mobile computing systems and tools**.

Additional Key Words and Phrases: Participatory Sensing; Physical Crowdsourcing; Mobile Crowdsourcing

ACM Reference Format:

Kapil Garg, Yongsung Kim, Darren GerGLE, and Haoqi Zhang. 2019. 4X: A Hybrid Approach for Scaffolding Data Collection and Interest in Low-Effort Participatory Sensing. *Proc. ACM Hum.-Comput. Interact.* 3, CSCW, Article 90 (November 2019), 28 pages. <https://doi.org/10.1145/3359192>

1 INTRODUCTION

Participatory sensing has developed into an effective method for actively engaging large numbers of people to report data about the physical world to help us understand it and to enable new services [37, 43, 52]. For example, birding hobbyists record their observations to help scientists track migration patterns [50]. Citizens in the U.S. make 3-1-1 calls to help city planners understand where city resources are needed [7, 34]. Users of mobile services such as Google Maps and Foursquare

Authors’ addresses: Kapil Garg, Northwestern University, Evanston, IL, USA, kgarg@u.northwestern.edu; Yongsung Kim, Northwestern University, Evanston, IL, USA, yk@u.northwestern.edu; Darren GerGLE, Northwestern University, Evanston, IL, USA, dgergle@northwestern.edu; Haoqi Zhang, Northwestern University, Evanston, IL, USA, hq@northwestern.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-ART90 \$15.00

<https://doi.org/10.1145/3359192>

actively contribute data about places to help others make plans around accessibility, dietary needs, and family needs [1, 3, 6].

Despite successful applications, meeting both the needs of users who contribute data and the system's goals for data collection remains a critical challenge for participatory sensing. On one hand, addressing the needs of data contributors—such as the desire for physical data collection tasks to be minimally disruptive [56], personally relevant [10, 50], and generally of value to them [34, 45]—is necessary for engaging enough contributors to actively make contributions. On the other hand, achieving data collection goals—such as obtaining *high-fidelity data* with detailed information about objects or events of interest at fine enough temporal and spatial scales—is necessary for understanding certain phenomena of interest and ensuring the usefulness of services that depend on such data [10, 17].

All participatory sensing systems must achieve some kind of balance between the needs of users and the system's data collection goals if they are to remain viable, but existing approaches are limited in how they strike this balance. *Opportunistic approaches*, where people are asked to actively contribute data along their existing routines when it is convenient for them to do so [50, 58], can better meet the needs of data contributors, but the opportunistic nature of the contributions makes it difficult to meet specific data needs. Contributors may not frequent certain locations, which makes it difficult to ensure high data coverage across locations and to keep dynamic data fresh. Even for locations that people frequently visit or pass by, the effort to contribute high-fidelity data can still be prohibitively high (e.g., filling out a full survey while walking around) [7, 34]. Low-effort opportunistic approaches can attract more contributors (e.g., [53, 56]), but using a low-effort approach has typically resulted in settling for low-fidelity data.

Directed approaches, where people are asked to fulfill a specific task that is not necessarily in their immediate vicinity, can be used to target specific system goals such as increasing data fidelity and coverage wherever it is needed. However, with a directed approach, meeting the needs of data contributors is difficult as it requires them to deviate from their existing routines [32, 47]. This disruption is offset either with monetary incentives [2, 4, 9, 28, 55], which may be infeasible or cost prohibitive in many domains, or through dedicated volunteers [15, 54], who may be difficult to recruit at scale in many domains.

To overcome the shortcomings of existing approaches, this paper introduces a hybrid, multi-stage approach to participatory sensing, *which integrates opportunistic and directed approaches by initially collecting data opportunistically and then using the collected data itself as a potential incentive for directed data collection*. A core idea is that collected data—even if low-fidelity at first—may provide useful information to data contributors that inspires further contributions in settings where the data itself is valuable to the user. For example, notifying people interested in events such as free food or guest lectures may inspire some people to go; while there, they may be asked to contribute additional information about what food is available and what the lecture is about. As another example, we may notify a hobbyist in a citizen science project such as eBird [50] of sightings of specific bird species of personal interest nearby, which may inspire them to go and to report additional information. We hypothesize that this approach can effectively promote additional contributions without additional disruption, while better meeting system goals for data such as increased fidelity and throughput.

To realize this approach, this paper contributes a new conceptual framework for data collection called 4X (eXplore, eXpand, eXploit, eXterminate) that provides a systematic process for leveraging both low-effort contributions at locations users pass opportunistically and contributions from users who are motivated to go to places outside of their routine when presented with personally-relevant,

community-collected data.¹ 4X simultaneously promotes user interests and builds towards system goals by first scaffolding low-effort, low-fidelity opportunistic contributions along users' routes, and then using collected data to draw other users to areas where the data and their interests align. In this manner, we explore how *the “cost” of directed approaches can be offset with the “value” of the information provided to the user and the event itself for those interested.*

To demonstrate and evaluate the effectiveness of the 4X framework, we implemented it in *LES* (Low-Effort Sensing), a low-effort sensing application on iPhone and Apple Watch that we designed to collect dynamically changing information about places and events around college campuses, such as coffee shops, libraries, and free food events. Through two user studies of *LES* ($N = 95$, $N = 18$), we demonstrate the advantages of 4X over opportunistic and directed approaches (Study 1), and the extent to which collected data can be used to promote additional data contributions from interested users who go out of their way (Study 2). Results from Study 1 show that 4X used collected data to create 34% more data collection opportunities without increasing reported disruption over an opportunistic approach; results further show that 4X is significantly less disruptive than a directed approach that notifies users of tasks at a distance regardless of whether there is data of interest to the user. Results from Study 2 show that 4X yielded 49% more data by directing users to locations of interest where they made additional contributions en route to and at target locations. These study results provide evidence for the effectiveness and potential of dynamic data collection processes for participatory sensing that draw on multiple sources of motivation for data, and that can better meet data collection goals as data becomes available and is used to induce further data contributions.

2 BACKGROUND

We are interested in advancing participatory sensing approaches that engage people to *actively contribute data about physical locations that they are in or are willing to go to* [14, 23, 30, 36, 37, 43, 52]. This is in contrast to prior work on machine-sensor based participatory sensing approaches that passively collect data from mobile device sensors or other custom sensors attached to people or objects (e.g., [8, 10, 11, 13, 19, 22, 57]).² While these prior systems can sense some attributes about the physical world, machine sensors are limited in what they can measure without active human participation.

More recent work involves humans-in-the-loop by using machine sensors to collect data and using remote crowd workers to analyze the data to derive useful information. This approach allows for sensing a wider range of phenomena, but is still limited by what machine sensors can observe. For example, Project Sidewalk asks online crowd workers to label Google Street View images with respect to accessibility issues present, like missing sidewalk ramps [48]. Though contributors to this project can help to interpret the collected data, this approach cannot easily capture dynamically changing information since the collected data is often stale by the time of analysis (e.g., as happens with seating availability at a coffee shop). Recent work on Zensors [38] is able to provide dynamic information about a location by sending a live camera feed to crowd workers and machine vision algorithms in real-time. However, this approach is still limited by machine vision algorithms that can only identify a narrow range of phenomena, and crowd workers who have limited access to context and who cannot move across the physical space where the phenomenon is occurring.

¹Our framework is inspired by and named after the 4X game design framework, popularized by turn-based strategy games such as Civilization.

²This passive machine sensing approach is sometimes referred to as “opportunistic sensing” in the literature [14, 23, 36, 37, 43]. However, this should not be confused with our use of the term, “opportunistic data collection approach,” that describes an active, participatory sensing approach in which people actively contribute data along their existing routines when it is convenient for them.

Given the limitations inherent in using machine sensors, we focus in this paper on better ways to actively engage people in participatory sensing. In what follows, we highlight the limitations of existing active participatory sensing approaches that restrict their data collection process to a single data collection mechanism, and argue for the need for hybrid, multi-stage approaches such as 4X to overcome the limitations of existing opportunistic and directed approaches.

2.1 Existing Approaches to Participatory Sensing: Opportunistic and Directed

To engage people in active data collection, *opportunistic data collection* approaches support users who provide data when it is convenient for them to do so, as part of their existing routines or through activities they are interested in. Example initiatives engage citizens and hobbyists to report infrastructural issues such as potholes, graffiti and broken street lights in the community [7, 34], to track the presence of various bird species [50], and to answer questions about locations people are at, such as the amount of time to get through airport security [42]. While these examples do not require users to travel out of their routine, engaging volunteers to provide useful data still requires designing for the interests and goals of data contributors [58]. System designers must balance the need for collecting high-fidelity data that is valid—which may require more stringent data collection protocols and thus more effort and interest on the part of contributors—with the competing goal of recruiting many participants to advance coverage [17]. While successful initiatives such as eBird [50] have garnered large-scale use and resulted in data collected from many contributors, recruiting enough willing contributors to projects is generally still a challenge [56].

To promote convenient contributions from a wider base of casual users, *low-effort opportunistic data collection* approaches introduce lightweight interaction techniques that seek to minimize the effort required to contribute. One approach presents tasks during smartphone unlocks to, for example, collect coarse-grained census data [56] and answers to microtasks [53]. Other approaches infer environmental data through immersive interactions embedded into a user’s habit-building practice (e.g., while going for a run) [40], or ask users to complete small tasks (e.g., look for a lost item; pick up and deliver a package) along their route while they are on-the-go [32, 33]. While these approaches may be useful for recruiting more contributors and increasing data throughput, they are generally limited to collecting low-fidelity data when high-fidelity data might be desired. Moreover, these approaches cannot direct users to help fill gaps in data coverage since data is only collected opportunistically along one’s existing routine.

In contrast to opportunistic approaches, *directed data collection* approaches actively direct users out of their routine towards tasks and areas where data is needed to target specific gaps in coverage or fidelity. Due to the deviation required from a user’s routine, directed approaches require high incentives which some prior works address by providing monetary rewards in return for a data contribution [2, 4, 9, 28, 55]. While paying participants allows for rich forms of data collection, the costs may be prohibitive for scaling many services, particularly those that require active monitoring or tracking dynamically changing information (e.g., campus events; the state of city infrastructure).

In the absence of monetary rewards, other directed approaches use gamification to align data goals with game mechanics; they direct users to collect useful data and travel as a byproduct of gameplay. For instance, PhotoCity recruited users to take photos of target areas from specific angles in order to make 3D models of buildings [54]. A contemporary example is Niantic’s Ingress Prime, which pits players against rival factions to create an immersive game environment in which dedicated gamers travel out of their existing routes to take in-game actions that produce crowdsourced data, such as walking routes, as a byproduct [5]. While effective for some use cases, designing appealing game mechanics that address a wide range of desired data collection goals remains challenging.

2.2 A Conceptual Framework for A Hybrid Approach to Participatory Sensing

Prior systems either opportunistically or directly ask users to complete tasks based on their convenience [53, 56], domain interests [50], or system’s data collection needs [9, 29], and in doing so, unnecessarily restrict their data collection process to a single data collection mechanism. Our work on 4X, in contrast, leverages the benefits of both opportunistic and directed approaches, allowing for the collection process to adapt based on the state of available data and user interest. This enables the collection of high-fidelity and high-coverage data while being minimally disruptive to users and providing them data of interest. Specifically, our approach considers ways to strategically gather initial pieces of dynamically changing, location-specific data that is used to attract users out of their way to locations of interest and inspire further data contributions, while other works focus on only using pre-existing data to recommend locations [16]. Unlike existing opportunistic data collection approaches that stop at using collected data as feedback to users when they contribute data (e.g., to see how their responses align with the rest of the community [56]; to know which nearby birds they have tracked so far [50]; to see summaries of collected ESM data to increase compliance [26]), 4X actively uses data collected by some users to provide relevant information of interest to other users, so as to promote further contributions. For instance, 4X may notify users about an event of interest (e.g., free food; or the sighting of a bird species) or relevant conditions of interest (e.g., an available table by the power outlets at the coffee shop) to draw them to locations where data gaps exist but the currently collected data still aligns with their interests. In this way, 4X implements a directed data collection approach that offsets *the “cost” of deviation with the “value” of the information provided to the users.*

Multiple prior works have explored technical frameworks for building context-aware and participatory sensing systems. Sensr [31], AWARE [20], Ohmage [51], and PartS [41] provide end-to-end frameworks for building general purpose participatory sensing platforms for collecting, managing, and analyzing prompted self-reports and sensor data streams, and for enabling researchers to run user studies with the platforms. In contrast to these technical frameworks for implementing data collection strategies, our contribution is a novel *conceptual framework for designing a multi-stage data collection process that becomes more directed as data of interest is collected.* 4X is a necessary complement to these existing technical frameworks because building a multi-stage data collection protocol requires making non-trivial design decisions about how and when to enact certain data collection mechanisms depending on where people are, what data is available, and what people care about that other works have not considered or evaluated.

3 4X FRAMEWORK

We propose a hybrid, multi-stage data collection approach for participatory sensing called 4X (eXplore, eXpand, eXploit, eXterminate) that collects low-effort, low-fidelity data opportunistically, and then uses these data to direct users to locations of interest to make additional contributions that build data fidelity and coverage; see Figure 1. Each stage of 4X aims to create more opportunities for data collection than opportunistic approaches without causing significant disruption to the user as directed approaches might, allowing for the building of fidelity and coverage. In this section, we describe the stages of the 4X framework and highlight key design decisions associated with implementing each stage.³

³We use the term “stages” to refer to the various data collection mechanisms that 4X uses as data becomes available and is collected, but do not mean to imply that only a single stage can occur at a time. For example, given data collected thus far, one user may contribute additional data opportunistically (via eXplore) while another is drawn from their routine to the location to contribute data (via eXpand). So while it may be useful to think about the data collection process as proceeding in stages, the 4X framework does not preclude the possibility of multiple stages occurring at the same time.

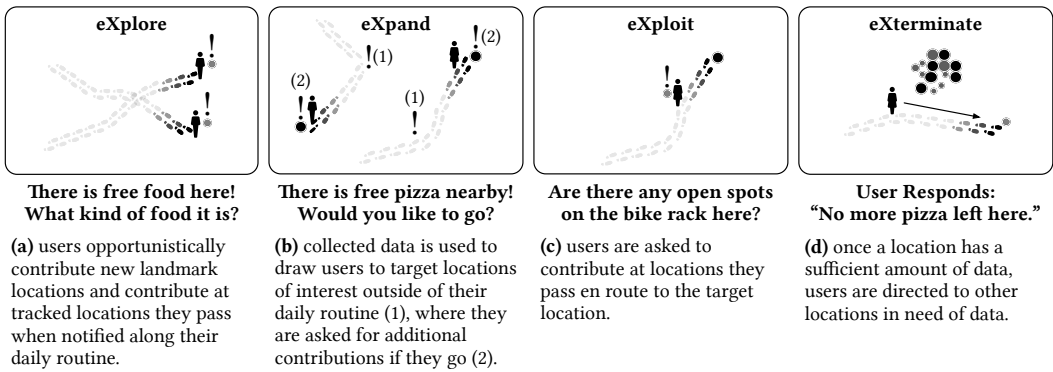


Fig. 1. The 4X framework scaffolds data in four stages. People first contribute opportunistically, marking new landmarks for tracking and responding when queried for information as they pass existing landmarks (eXplore). Collected data then draws people to target locations outside of their path, where they can be queried for additional contributions (eXpand). People may also be queried en route to the target location for contributions (eXploit). As data at a location fills (indicated by the size and the darkness of circles), users are directed to other regions and locations in need of contributions (eXterminate).

3.1 eXplore: Collecting Opportunistic Contributions

The eXplore stage (Figure 1a) allows users to contribute opportunistically, both as a primary responder to indicate the presence of an event or location (e.g., there's someone giving away free food here) and as a secondary responder to contribute additional data when notified *at-location* during their daily routine (e.g., there is someone giving away free food here—can you tell us what kind of food it is?). Primary responder contributions act as landmarks, allowing for more data and interest to build up around specific locations and events. Subsequent contributions from secondary responders scaffold the data collection to increase data fidelity.

During the eXplore stage, contributions are collected opportunistically from those who can contribute conveniently along their existing routine. To limit disruption to users' existing routines, participatory sensing applications that instantiate the eXplore stage should make it easy for users to make low-effort contributions. This can be supported by low-effort interaction techniques for primary and secondary responders that allow contributions to be made in seconds (e.g., [53, 56]). Collected low-fidelity contributions from individual users can then be scaffolded and combined to build data fidelity across multiple contributions and users. Together, these techniques allow individual users to contribute conveniently while still building the needed data fidelity to meet system goals, so that the eXplore stage (and 4X more generally) can collect convenient contributions from larger groups of casual users rather than relying on small groups of dedicated volunteers.

3.2 eXpand: Directing Users to Places of Interest

The eXpand stage (Figure 1b) uses *collected data as an incentive to explicitly pull users away from their routine* to nearby locations of interest. When the collected data aligns with a user's interest, eXpand sends an *at-distance notification* (e.g., there is free food in a nearby building, would you like to go there?) in order to build data fidelity at the *target location of interest*. Subsequently collected data can then be used to draw in other interested users. In this way, instead of only relying on opportunistic contributions from eXplore to build data fidelity, eXpand seeks additional, directed contributions from those who do not immediately pass by target locations.

Unlike directed approaches that indiscriminately notify users to make contributions outside of their routine, 4X *selectively notifies* users about nearby locations only when data of interest is available to motivate them to deviate from their routine. Providing users with data that they find valuable helps to offset the cost of deviating from their routine, which limits perceived disruption since users value the information and may wish to go out of their way to the location of interest. Through selective notifications, eXpand creates new contribution opportunities as a byproduct of surfacing existing data of interest back to the user while limiting the perceived disruption since users are receiving information of interest.

Implementing the eXpand stage of 4X into a participatory sensing application requires (1) building a *user model* that captures individual users' interest in the collected data, and their willingness to go out of their way; (2) setting a *notification selection criteria* that determines whom and when to notify a user based on their interests (i.e., their user model); and (3) setting a *question selection criteria* to determine what additional information to solicit from users when they do decide to visit locations upon being notified. We discuss each of these components below:

- (1) The **user model** determines how likely a user is to go out of their way when presented with information about locations and events of interest to them. A user model may consider (a) the kind of location (e.g., free food events; coffee shops) and the types of data about a location (e.g., food type; seating availability) that a user is interested in; (b) the amount or specificity of available information for each kind of location needed to draw in a user (e.g., private seating versus private seating by the outlets at a coffee shop); and (c) any contextual factors that may affect a user's likelihood to deviate from their routine, such as their schedule or how far they would be willing to go out of their way to visit a location of interest. To create a user model, a 4X system designer may ask users to directly express what information they find valuable and what they would want to be notified about, or train a model using machine learning to predict how likely a user is to go out of their way based on observations of their past decisions to act (or not) on data of interest presented to them.
- (2) The **notification selection criteria** is used to decide whether to notify a user about a location or event of potential interest. Based on the collected data and the user model, this criteria considers the likelihood of users to go out of their way and determines how much available information is sufficient to notify a user, and at what distance. While it is generally advisable to notify users with information that is interesting or valuable to them, the notification selection criteria can be conservative and specify only sending information that is extremely valuable to the user (e.g., free food that a user really likes), or be aggressive and specify sending all information of potential interest. Likewise, the criteria may specify sending information of interest to users only when they are a short walk from the location, or sending information when a much larger deviation is needed. These decisions broadly affect how many data collection opportunities are made available, and how valuable or disruptive users may find the notifications sent via eXpand.
- (3) The **question selection criteria** is used to decide what piece of information to solicit from a user when they do decide to go to a target location of interest that is suggested by eXpand. The question selection criteria can be used to prioritize gathering information that is valuable to users generally (e.g., more users want to know if there is seating near power outlets than near windows), which provides direct value to those interested and opens up future data collection opportunities from those who visit the location. The criteria can also be used to prioritize data collection goals that are valued by the system or that better meet the needs of a particular subset of the users (e.g., collecting information about accessibility features).

3.3 eXploit: Creating Contribution Opportunities En Route

The eXploit stage increases data coverage by collecting data at places that a user now passes *en route* to a nearby eXpand target location and *around the target location* itself that a user would previously not have gone to (see Figure 1c). Like eXpand, opportunities for additional contribution in eXploit are presented as a byproduct of the user choosing to go to a target location. As these contributions are collected opportunistically like eXplore, the design decisions discussed there to make contributions convenient and low-effort similarly apply here.

While en route to a location, eXploit can request contributions from a user about tracked locations they pass to increase data coverage in regions between the user's original location and their target location. Since the effort to make a contribution is low and since users are already going to a location of interest based on information that the system provided to them, we expect that asking for these additional contributions will not be perceived as disruptive by the users.

After a user has reached the target location, eXploit can also ask users to make contributions in the region around the target location to further build data coverage. For example, birders contributing to the eBird [50] citizen science initiative may be willing to contribute information about other birds in the area after being notified about a bird they are interested in seeing at a nearby location. In this situation, it is advisable to have the contribution opportunities align with things that users are interested in contributing to, since some (small) deviation may still be involved.

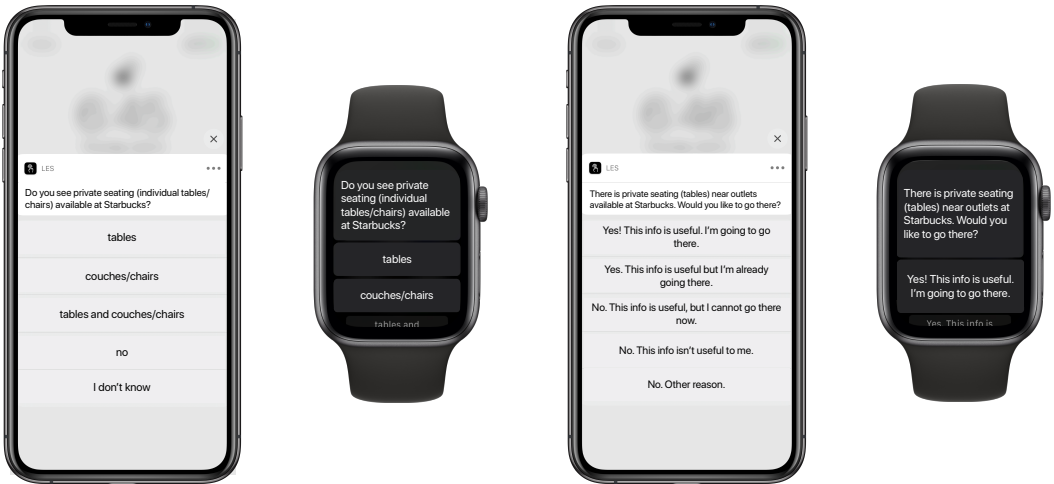
In summary, we hypothesize that through the eXpand and eXploit stages, 4X increases the number of contribution opportunities and the actual number of contributions relative to the eXplore stage alone, without increasing disruption since eXpand only occurs when data of interest is available and the interactions for data contribution during the eXploit stage remain low-effort. Together, these two stages can increase both data fidelity and coverage, as users contribute data at, en route to, and around specific locations of interest.

3.4 eXterminate: Shifting Focus to New Regions and Locations

Finally, the eXterminate stage is used to determine if tracking for a location should restart or cease based on the current state of the collected data (see Figure 1d). We consider two different conditions for eXterminate: *data staleness* and *data fidelity thresholds*.

When eXterminate detects that data at a certain location has become stale and may no longer be valid, data scaffolds are cleared and data collection at the location begins anew. In cases where users want to know about dynamically changing data at different locations, eXterminating when the data is stale helps 4X to ensure that users are provided with up-to-date information that may motivate them to deviate from their routine. To determine that data has become stale, a participatory sensing system may request information from users to help determine whether data has become invalid. For example, a system tracking free food events may ask "Is there still food available?" to determine whether to continue tracking.

If instead eXterminate identifies that the collected data has reached a data fidelity threshold for a location (i.e., most or all of the desired information has already been collected), tracking at the location stops in order to promote data collection efforts at other locations. This helps a 4X system to work towards addressing other pressing data needs rather than over-emphasizing data collection efforts at places where they are no longer needed. To implement such a threshold, applications may consider including a flag in the data scaffold which indicates to eXterminate that enough data at a location for a certain time period has been collected. A simple example of this is checking to see if all questions in the data scaffold have already been answered by users.



(a) Queries for additional data contributions contain any needed context to answer the question, and the possible categorical responses for the query.

(b) Notifications about nearby places of interest are sent to users when the available data aligns with their data interests (as reported in the pre-installation survey).

Fig. 2. LES interfaces for contributing data and for receiving data of interest. Users respond to simple queries about locations they are in or are about to pass (a), and receives valuable information about nearby places of interest that they care about (b). All responses can be made through contextual notifications on the device's lock screen without needing to open the installed application.

4 LES: CAMPUS DATA COLLECTION WITH 4X

Having detailed the 4X framework, we now present *LES* (Low-Effort Sensing), a location-based participatory sensing system built for the iPhone and the Apple Watch that instantiates the 4X framework to collect dynamic information about local places and events around college campuses from mobile crowds. In this section, our goal is to illustrate how the conceptual 4X framework might be implemented into a practical application. We describe (1) what *LES* is; (2) how *LES* instantiates each stage of the 4X framework; and (3) the technical implementation of *LES*.

4.1 What is LES?

LES is an application for sharing dynamic information about places and events on college campuses. With *LES*, students can contribute information about coffee shops, libraries, gyms, and free food events such as seating or equipment availability and what kind of free food is being given out. Users are notified and asked to contribute data about tracked locations as they are about to pass the location using a low-effort interaction that can be performed from the lock screen of their phone. They can also receive information of interest to them about nearby locations (e.g., open seating by the power outlets at their favorite coffee shop) when it is available. Unlike prior work on *TASKer* [28, 29] that studied collecting information about campus events by paying students to make contributions, *LES* is designed to collect convenient, low-effort contributions on users' existing routes, and to use people's interest in events and situations (e.g., free ice cream on a hot day; open seats at a coffee shop) to promote contributions at places outside of their current route.

4.2 Instantiating 4X in LES

In this section, we detail how we instantiate each stage of the 4X framework in *LES* for collecting data about locations and events on college campuses.

4.2.1 eXplore in LES. To instantiate the eXplore stage of 4X, LES notifies users to make opportunistic contributions when they are about to walk past locations tracked by the application. To limit disruption when making these contributions, LES implements contextual notifications that appear on the device’s lock screen and can be responded to without needing to unlock the device or open the application when the user is about to pass by a tracked location; see Figure 2. Each notification contains the relevant contextual information derived through earlier responses by other LES users, a query for additional information about the location in the notification body, and the categorical responses to answer the query when the notification is opened; see Figure 2a. In addition to the relevant categorical responses for the query, we include the “I don’t know” response as a way for users to say that they received the notification, but were unable to answer it; this response was included based on early user testing where users said they wanted a way to still respond to the notification if they could not answer, instead of simply ignoring it. Such contribution techniques have been proven to be low-effort in prior works such as Slide to X [53] and Twitch [56].

To create higher-fidelity data from low-effort contributions, LES uses *information scaffolds* that deconstruct the desired, richer information into low-fidelity components with corresponding questions that can be easily responded to with low-effort contributions. LES includes data scaffolds for a variety of locations on and around college campuses, including libraries, gyms, coffee shops, and free food events. These locations and their associated questions were informed by needfinding surveys of target users where respondents indicated what places around the campus they would like to know information about, and what information they would like to know. Rather than only capturing coarse information about the tracked locations (e.g., that private seating at a coffee shop is available), we chose to have our information scaffolds capture more detailed information (e.g., that private seating near the windows at a coffee shop is available) that could be useful for meeting the different information needs of users during the eXpand stage.

As an example, LES would use the following scaffold and question breakdown for collecting data about coffee shops with the question ordering being 1, 1a, 1b, etc.

There is *private seating* [*tables/couches/chairs*] near [*outlets, windows*] and/or *shared seating* [*communal tables*] near [*outlets, windows*] at Starbucks.

- (1) Do you see *private seating* (*individual tables/chairs*) available at Starbucks? [Tables, Couches/Chairs, Tables and Couches/Chairs, No]
 - (a) Are any of these near outlets? [Yes, No]
 - (b) Are any of these near the windows? [Yes, No]
- (2) Do you see *shared seating* (*communal tables*) available at Starbucks? [Yes, No]
 - (a) Are any of these near outlets? [Yes, No]
 - (b) Are any of these near the windows? [Yes, No]

If any LES user responded “No” to either questions 1 or 2, then the following sub-questions would not be asked to other users. When a user is presented with a notification, all information is included in the notification to help direct them to what specific information is needed by the system. For example if there were private tables available, but not near outlets, a user passing the tracked Starbucks would be asked, “There is private seating (tables) available at Starbucks. Can you tell us if there are any near the windows?”

4.2.2 eXpand in LES. Once LES has collected some data opportunistically, it begins the eXpand stage by selectively notifying users when their interests align with the collected data. LES does this in two phases: it first sends an initial notification to users letting them know of information about a nearby location of interest and asks if they would like to go to the location (e.g., “There is private seating (tables) near outlets available at Starbucks. Would you like to go there?”; see Figure 2b), and then sends a second notification to users who do decide to go to the target location to request

additional data there (e.g., “There is private seating (tables) available at Starbucks. Can you tell us if there are any near the windows?”).

As discussed in the previous section, implementing the eXpand stage of 4X requires forming a user model, setting a notification selection criteria, and setting a question selection criteria. We discuss how we implemented these components in LES and the associated design decisions below:

- (1) **User model:** LES implements a simple model of user interest that focuses on what data users find valuable (e.g., a user wants to know about private seating at a coffee shop). To develop this model, we first conducted needfinding to determine the kinds of campus locations and types of data about these locations that students are generally interested in. We then asked each LES user to fill out a pre-installation survey that asked them, for each type of data available for each kind of location, if they would like to receive notifications when such data is available (see each of the study sections for more details on the pre-installation survey).
- (2) **Notification selection criteria:** Once LES has collected some data from opportunistic contributions, it selectively notifies users if the collected data matches their user model (i.e., if the data matches what the user indicated what they wanted to be notified for in the pre-installation survey). Although 4X can notify users at various distances during eXpand based on level of their interest, we chose to notify nearby users who are within a set distance radius of the target location (e.g., 300 meters; see Technical Implementation for more details).
- (3) **Question selection criteria:** LES prioritizes collecting the types of data that users are more interested in knowing about; such information is more valuable to users, and also more likely to lead to additional data contribution opportunities from the larger group of users who may wish to go to the location should the collected information be of interest to them. To do this, LES rearranges the information scaffolds presented earlier to first ask questions that collect the data that most users of the system would be interested in knowing, based on the pre-installation surveys used to build the earlier user models. For instance, if we found that more users expressed a preference in knowing about shared seating rather than private seating, LES would ask question 2 and its sub-questions before asking question 1.

4.2.3 eXploit in LES. To implement the eXploit stage of 4X in LES, we consider how LES may notify and request additional contributions from users en route to a nearby eXpand target location. To collect additional data en route, LES notifies users to make contributions using the same interaction technique as used for eXplore when they pass tracked en route locations. To provide users with opportunities to make contributions en route, LES includes additional locations like bike racks and parking lots that users may frequently encounter while on their way to an eXpand target location.

While eXploit can also be used for requesting contributions around a location of interest that a user is going to, we chose not to implement this use case in LES because (a) the locations of interest were mostly spread out across the university campus; and (b) in our domain, users interested in one location did not often have an interest in visiting nearby locations.

4.2.4 eXterminate in LES. To instantiate the use of eXterminate to handle situations when data has become stale, LES uses *timed refresh cycles* and *verification questions* to signal when data scaffolds are no longer valid and should be cleared. Timed refresh cycles act as the base reset condition for data scaffolds to ensure that the presented data is still correct and up-to-date, and trigger based on how long the current data has been in the scaffold. If any data has been in the scaffold for longer than a specified threshold time, the scaffold is fully cleared. Then depending on the type of location, tracking of the location will either restart at places that continue to generate new data (e.g., coffee shops; libraries) or cease for events that are no longer happening (e.g., free food events). When no further questions can be asked from a data scaffold, LES uses a verification question to see if the

current data is still correct. If a user responds that the data is no longer correct (e.g., no more free food left), LES clears the data scaffold and either restarts or stops tracking similarly to the timed refresh cycle.

While eXterminate may also use thresholds for data fidelity to shift data collection focus to other locations, we chose not to implement this feature in LES since there was not a need to promote or distribute data collection at all locations evenly. In other words, we let LES collect data at places where people naturally passed and where people were interested in knowing about the data.

4.3 Technical Implementation

LES consists of a client application written in Swift for iOS and Watch OS, and a back-end built using Node.js and MongoDB. The iOS front-end handles real-time outdoor and indoor location tracking using geolocation and Estimote Bluetooth iBeacons respectively, and the generation of contextual notifications using Apple's UserNotifications framework. When near a tracked region, the front-end presents users with a notification containing the current information about the location, the query for the next piece of information, and the possible answers to the query (see Figure 2a); similarly, the front-end will notify users about nearby places of interest within the distance threshold specified by the notification selection criteria (see Figure 2b). Responses to notifications are sent to the back-end that handles all tasks related to building data scaffolds, generating notification contents and queries, and syncing information with all users' applications in real-time.

As implemented, LES sets a notification radius of 300 meters (approximately 1.5 blocks in the deployed city) for eXpand, so that the radius will be large enough to require users to make some deviation from their routine, but not so far that the users would never go out of their way. In order to avoid spamming users with multiple notifications when they pass through areas with multiple tracked locations in close proximity, we additionally set a 10 minute notification interval between notifications so that once notified about a data collection opportunity via eXplore or eXpand, the user will not receive another notification within the threshold.

5 STUDY 1: COMPARING 4X TO OPPORTUNISTIC AND DIRECTED APPROACHES

We present in the following sections two user studies of LES that demonstrate the advantages of the hybrid, multi-stage 4X framework over purely opportunistic and directed data collection approaches (Study 1), and the extent to which collected data can be used by 4X to promote additional data contributions from interested users who go out of their way (Study 2). In both studies, we are interested in understanding how the process of providing people with specific information that is of interest to them as it is dynamically acquired can better meet data collection goals while avoiding unnecessary disruption to users.

In Study 1, we compare 4X to an Opportunistic data collection approach and a Directed data collection approach to study how 4X might better meet the needs of users (e.g., convenience and low disruption) than a Directed approach, and provide more opportunities to meet the needs of the system (e.g., data collection goals) than an Opportunistic approach. Specifically, our Opportunistic approach aims to meet user needs by only asking for contributions when users can conveniently contribute along their existing routine. We expect that this will lead to some collected data, but also miss opportunities to collect more data when people would have been willing to visit locations outside of their routine. Meanwhile, our Directed approach aims to meet system data collection goals by asking users to contribute data at any tracked location they might be nearby, which they may or may not pass as part of their existing routine. This approach can potentially lead to more contributions from a larger set of users, but at the cost of greater disruption due to more notifications and deviation from users' routines. We expect that 4X will provide users with more data contribution opportunities than the Opportunistic approach by also notifying them about

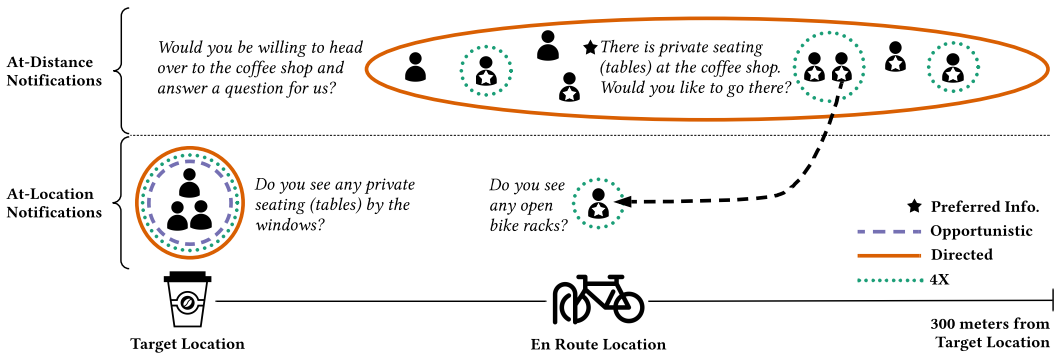


Fig. 3. Illustration of the notification policies for the Opportunistic, Directed, and 4X conditions in Study 1. In all conditions, users receive at-location notifications that ask them to contribute additional information when they are about to pass by a tracked task location. Users in the Directed and 4X conditions also receive at-distance notifications up to 300 meters away from the task location. The Directed condition sent these notifications regardless of whether preferred information is available (information is included when available), whereas the 4X condition only sent these notifications if preferred information is available. A star on a person indicates that preferred information of interest is available for that user at the time of notification. Finally, 4X also asks users to make contributions en route when they decide to go to the target location.

nearby tracked locations, but will be less disruptive than the Directed approach, since requests to travel to locations outside of their routines are only made when collected data suggests that they may be interested in knowing. To summarize, we hypothesize that:

- H1:** 4X creates more opportunities for data collection over an Opportunistic data collection approach without increasing disruption.
- H2:** 4X is less disruptive than a Directed data collection approach and sends significantly fewer notifications.

5.1 Method and Analysis

5.1.1 Participants. We recruited 95 undergraduate and graduate students of a mid-sized U.S. university through mailing lists, social media, and word of mouth. Participant ages ranged from 18 to 28 ($M = 20.30$, $SD = 1.93$), with 72 female and 23 male participants. The study took place over 14 days, during which participants completed a pre-study survey, used LES as a part of their daily lives, and completed a post-study survey. We compensated participants \$20 for their time spent on surveys and installing LES, but did not incentivize their behavior during the study (i.e., no monetary incentive was provided for task completion).

5.1.2 Procedure. The 95 study participants were randomly assigned to one of three study conditions: Opportunistic (32), Directed (32), and 4X (31).⁴ We chose a between-subjects design instead of a within-subjects design to: (1) avoid any carryover effects on our measures when users switched between conditions since each condition only had subtle, hard-to-notice differences in the notification policy; (2) control and mitigate weekly variability in routines; and to (3) avoid study fatigue associated with a necessarily longer within-subjects study.

⁴We removed 1 participant from the 4X condition because they received far more notifications than should have been allowed due to a temporary technical malfunction. This left us with log data from 94 participants (30 for 4X; 32 each for Opportunistic and Directed). 9 of these users did not fill out the post-study survey (2 for 4X; 3 for Opportunistic; and 4 for Directed), leaving us with survey responses from 86 participants (29 each for Opportunistic and 4X, and 28 for Directed).

Figure 3 summarizes the notification policies of each condition, detailing when users would receive notifications to contribute additional information based on the condition they are in. In all three conditions, users receive *at-location* notifications as they pass by landmark locations (eXplore). Users in the 4X condition additionally receive *at-distance* notifications that ask them to go to landmark locations when there is data of interest about the location available (eXpand); they are also asked to contribute data about other locations *en route* should they decide to go (eXploit). Users in the Directed condition receive at-distance notifications similar to 4X if data of interest is available, but would also receive a generic contribution notification (e.g., “We need some information about the Starbucks nearby. Would you be willing to head over and answer a question for us?”) anytime they were near a landmark location with no data of interest available. In both the 4X and Directed conditions, all at-distance notifications are sent when users are within 300 meters of a landmark location. In all conditions, outside of the notification interval set by LES to avoid notification spamming, notifications are sent based on the notification policies of the 4X framework and the data collection approaches used in the Opportunistic and Directed control conditions.

We pre-populated each instance of LES with seven landmark locations around campus and the surrounding area (three coffee shops, two workspaces, and two gyms). We selected categories and locations with dynamically changing status that students were interested in knowing about (e.g., when tables or gym equipment become available) based on previous needfinding of university students. In addition, one of the authors hosted free food events around campus every other day at various times over the course of the study, which served as additional landmark locations. Data scaffolds for each of these locations start out empty, and can go up to five levels deep as users contribute reports to build up data fidelity. Scaffolds were cleared of information either after four hours, or when users reported that the data is no longer correct (eXterminate).⁵ Beyond these landmark locations, we added six locations of bike racks and parking lots where users in the 4X condition were asked if free spaces were available when they passed by en route.⁶

To present information of interest to users, we collected through a pre-study survey each users’ high-level preferences over the kind of information they were interested in being notified about. For example, a participant who is generally interested in private tables near windows at a coffee shop but not in communal tables can specify that they are interested in private tables, and in sitting by the window. Whenever preferred information is available, it is included in any notifications that users in the 4X and Directed conditions receive. For example, a user interested in coffee shop seating may receive a notification at-distance that reads: “There is private seating (tables) near outlets at Starbucks. Would you like to go there?” Users in the Opportunistic condition were not notified of information they may care about, but can access such information via a “For You” page within LES; this page was also available for 4X and Directed condition users. In cases where no preferred information is available, the Directed condition simply asked users if they would be willing to head to a nearby landmark location to contribute some information; see Figure 3.

We used user responses to at-distance notifications as an ecological momentary assessment (EMA) [49] to assess (a) whether they found the information presented to them useful; and (b) whether they were going out of their way based on the information presented or if they just happen to be already going to a landmark location; see Figure 2b. Users can also specify reasons for not going, such as having a scheduled event.

Beyond EMAs, users completed a post-study questionnaire following the 14 day usage period, which asked them to reflect on the disruption and value of LES during their usage, recall times

⁵Four hours was used as the refresh time since we could not account for scheduling and mobility patterns of users in this study, and we wanted enough time for data to build in fidelity so that it could be used in the eXpand stage.

⁶Per day, a user may be asked to contribute information at multiple locations, but never more than once per location (within a data refresh cycle) and at only one location en route to avoid over-notifying users.

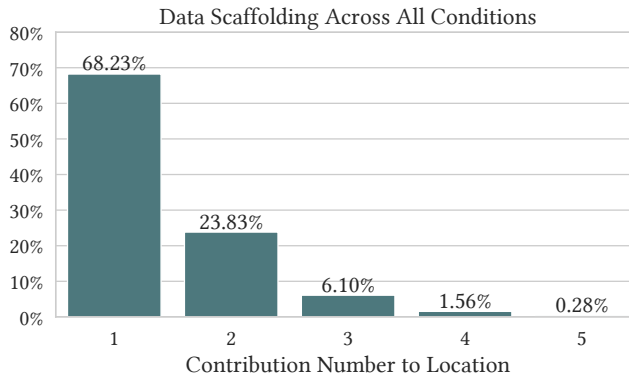


Fig. 4. Percentage of users' contributions that built up data fidelity by scaffolding together multiple contributions beyond the first. 31.77% of all contributions made by users increased data fidelity beyond the first contribution (e.g., private tables available at a coffee shop), leading to higher-fidelity data from the scaffolded low-fidelity contributions (e.g., private tables near outlets available at a coffee shop).

when they were asked for contributions and elaborate on why they did or did not respond, and recall times when they were notified to go to locations outside of their routine and reflect on why they may or may not have gone. We coded these open-ended survey responses along the dimensions of (1) ease of contribution; (2) the value users received in getting notified about data of interest; and (3) the disruptiveness of the notifications received. Codes were aggregated into counts for each condition to present overall trends related to these dimensions. We were interested in these particular dimensions as they directly map back to the key design goals of 4X, namely (1) broadening participation with low-effort contributions; (2) increasing data fidelity and coverage by drawing users from their routines with data of interest; and (3) limiting disruption when asking for users to deviate or make additional contributions.

5.1.3 Measures and Analysis. We measured how disruptive users found each condition by asking how frequently they felt disrupted by LES notifications over the two-week study period on a 5-point Likert Scale (1: Never, 5: Always) in the post-study survey. To show that 4X creates additional data collection opportunities, we compared the number of notifications sent to users in the 4X condition at-location and at-distance.

We used the EMAs and post-study surveys to better understand why users decided to contribute or not, both at-location and at-distance, across the three conditions. To track specific cases where users in the 4X and Directed conditions went out of their way and contributed additional data, we used the EMA responses and log data to determine, respectively, (a) people's decisions to go out of their way; and (b) whether they went to locations they were notified about and made additional data contributions. We count all user responses at-location as actual data contributions, with the exception of when they select the "I don't know" option.

5.2 Results

Results show that LES effectively scaffolded low-effort contributions to build higher-fidelity data. During the 14 day deployment, the 94 users in Study 1 made a total of 705 data contributions. Of these, 224 contributions (31.77%) were subsequent contributions after the initial that were used to build data fidelity. While the initial contributions informed users when seating became available at libraries and when free food was available, subsequent contributions allowed users to know that there was private seating near windows and that there was free pepperoni pizza. Figure 4 shows the

Table 1. Breakdown of contributions and notifications sent by condition in Study 1.

	Opportunistic	4X	Directed
Valid Contribution Origin			
At-Location	281	220	187
At-Distance (Willing to Deviate)	–	1	5
At-Distance (Already Going)	–	1	10
Total Valid Contributions	281	222	202
Notifications Sent			
At-Location	527	446	402
At-Distance	–	152	1059
Total Notifications Sent	527	598	1461

frequency at which user contributions built up data fidelity by integrating multiple contributions beyond the first.

Across all conditions, users were notified a total of 2586 times with an average number of notifications per user per day of 1.18 for Opportunistic, 1.42 for 4X, and 3.26 for Directed. Table 1 breaks down the number of notifications sent in each condition, both at-location and at-distance. Users responded with information (i.e., made a valid contribution) 51.27% of the time when they were notified at-location to contribute additional data (53.32% for Opportunistic; 49.78% for 4X; and 50.25% for Directed). In post-study surveys, users cited ease of interaction as their primary reason for contributing at-location. Across all conditions, 47 out of 86 users (54.65%) mentioned ease as the main reason for contributing while at-location. For example, P17 (4X) wrote: *“I responded because I was skateboarding by [coffee shop], and it wasn’t too hard to slow down, glance through the windows, and respond on my phone.”* Some users also mentioned contributing because they felt that their contribution would be helpful to others using LES, like P22 (Opportunistic): *“It was simple enough to respond, and I went into [coffee shop] to look for myself anyway. I figured it would be good to share this information with other people since I would want to know myself.”*

5.2.1 H1: 4X Creates Additional Contribution Opportunities Without Adding Disruption Over the Opportunistic Approach. Our results show that 4X created more data collection opportunities without adding disruption over the Opportunistic approach. 4X created 34.09% more data collection opportunities by presenting users with information they wished to be notified about; beyond the 446 notifications sent at-location (1.06 notifications per user per day on average), 4X sent an additional 152 notifications at-distance (0.36 per user per day on average); see Table 1. Despite sending more total notifications than the Opportunistic condition (598 vs. 527), users did not find 4X to be more disruptive; the average reported disruption was 2.08 (SD = 0.63) for 4X and 2.10 (SD = 0.61) for the Opportunistic condition. A Mann-Whitney U Test showed that the distributions of the reported disruption were not significantly different ($n_1 = 29$, $n_2 = 28$, $U = 412.000$, $p = 0.925$).

A plausible explanation for why users did not find 4X to be more disruptive than the Opportunistic condition despite the additional notifications is that the 4X users generally found LES to be more valuable to them than the Opportunistic users because they received notifications with information of interest to them. 12 out of 29 users (41.38%) in the 4X condition mentioned how they liked being presented with data of value to them, and some, like P21, indicated how it would help them decide where to go: *“I liked how easy it was to be helpful and the idea that I can check to see if there are free places for me to do work before I actually went to those places.”* In contrast, users in the Opportunistic

condition did not receive notifications with information of interest to them, which some users wished they did. For example, P26 said: *“I would have liked if the notifications I received were instead to inform me of something that I wanted rather than asking me for inputs. Especially if there was free food somewhere, I would have liked to receive a notification about it.”*

Despite creating additional contribution opportunities and presenting data of interest to users, in Study 1 only four users in the 4X condition deviated from their routes, which yielded only one additional data contribution. We identified three partial confounds that may have led to this outcome, which we will address later through Study 2. First, Study 1 was conducted during a period of the school year where many users had exams and other scheduling constraints, which limited the ability of users to go out of their way. In 35.53% of responses to at-distance notifications, 4X users indicated that scheduling constraints prevented them from visiting the location even though they found the information presented useful. Second, users experienced frigid and stormy weather conditions throughout the study period, which likely lessened their desire to go out of their way to any location. Third, from the post-study survey results, we found that in some cases the pre-study survey did not sufficiently capture users’ notification preferences and thus sent them information that was not valuable to them. For example, while the pre-study survey allowed users to specify that they are generally interested in coffee shops, they could not specify that they only cared for a particular coffee shop and preferred not to be notified about others. Examples such as this contributed to the 18.42% of notifications sent to 4X users who indicated that the information presented was not useful to them. In Study 2, we will show how removing these confounds and improving our notification preference survey led to a significant number of additional contributions from 4X users who go out of their way upon receiving at-distance notifications containing data of interest to them.

5.2.2 H2: 4X is Less Disruptive than the Directed Approach. Having demonstrated some of the potential advantages of using 4X over the Opportunistic approach, we turn to compare 4X to the Directed approach. We found a significant difference in reported disruption; the average reported disruption was lower for 4X ($M = 2.08$; $SD = 0.63$) than for the Directed condition ($M = 2.54$; $SD = 0.86$). A Mann-Whitney U Test showed that the distributions of reported disruption were significantly different ($n_1 = n_2 = 29$, $U = 280.500$, $p = 0.023$).

A primary cause for the increased disruption in the Directed condition is the number of additional notifications sent. The Directed condition sent users nearly 7 times more at-distance notifications than 4X did (1059 vs. 152) and 2.5 times more notifications overall (1461 vs. 598); see Table 1. On average, the users in the Directed condition received a total of 3.26 notifications per user per day versus 1.42 for the 4X condition. 16 out of 29 users in the Directed condition (55.17%) mentioned that they did not like the quantity of at-distance notifications they received without any data of personal interest. P5 noted: *“Realistically, I’m not going to go out of my way to answer a question. I know this is out of self-interest and it deters from everyone’s experience, but being honest.”* This suggests that while directed approaches can reach more users, in practice they might not lead to the kind of system that is desirable or sustainable due to the increased disruption and lack of value for attracting users to make additional data contributions without providing additional incentives.

6 STUDY 2: 4X YIELDS ADDITIONAL CONTRIBUTIONS

Study 1 demonstrated some of the potential advantages of 4X over Opportunistic and Directed approaches, namely that 4X can create additional data contribution opportunities without increasing disruption. However, due to potential confounds with respect to scheduling constraints, weather, and unexpressed preferences, only a few users went out of their way in response to notifications sent at-distance and contributed additional information. We designed Study 2 to address these

Notification Preferences	Likelihood to Go out of Your Way
<p>LES knows that...</p> <p>Would you like to be notified about Coffee Shops, given this much information?</p> <p>a nearby coffee shop has private seating (individual tables) available. <input type="radio"/> Yes <input checked="" type="radio"/> No</p> <p>a nearby coffee shop has private seating (individual tables) near outlets available. <input type="radio"/> Yes <input checked="" type="radio"/> No</p> <p>a nearby coffee shop has private seating (individual tables) near windows available. <input checked="" type="radio"/> Yes <input type="radio"/> No</p>	<p>LES notifies you that...</p> <p>How often would you go out of your way to a Coffee Shop, given this much information?</p> <p>a nearby coffee shop has private seating (individual tables) available. <input type="radio"/> Always <input type="radio"/> Sometimes <input type="radio"/> Rarely <input checked="" type="radio"/> Never</p> <p>a nearby coffee shop has private seating (individual tables) near outlets available. <input type="radio"/> Always <input type="radio"/> Sometimes <input checked="" type="radio"/> Rarely <input type="radio"/> Never</p> <p>a nearby coffee shop has private seating (individual tables) near windows available. <input type="radio"/> Always <input checked="" type="radio"/> Sometimes <input type="radio"/> Rarely <input type="radio"/> Never</p>
(a) Notification Preferences	(b) Interest Preferences

Fig. 5. Notification and interest preferences for Study 2. In this example, the participant wishes to be notified about private seating at coffee shops only when they are near windows (a), and reports a stronger interest in seating near windows than near outlets (b).

confounds, and additionally measure the extent to which collected data can be used to promote additional data contributions from interested users who go out of their way. Specifically, Study 2 provides evidence for the following hypotheses:

- H3:** 4X collects additional data from users who respond to at-distance notifications and go out of their way to visit places of interest, while still being minimally disruptive.
- H4:** Users are more likely to go out of their way when presented with information that is more valuable to them.

6.1 Method and Analysis

6.1.1 Participants. We recruited 18 undergraduate and graduate students of a mid-sized U.S. university through mailing lists, flyers, social media, and word of mouth. Participant ages ranged from 19 to 31 ($M = 23.93$, $SD = 3.69$), with 8 female and 9 male participants (one preferred not to specify). The study took place over 14 days, during which participants completed a pre-study survey, used LES as a part of their daily lives, and completed a post-study survey. We compensated participants \$30 for their time spent on surveys and installing LES, but did not incentivize their behavior during the study (i.e., no monetary incentive was provided for task completion).

6.1.2 Procedure. All 18 users were assigned to a single condition in which they use the 4X version of LES (identical to Study 1) for 14 days.⁷ We ran Study 2 during the university's summer session, when scheduling constraints are less restrictive and the weather is favorable compared to Study 1. In the absence of scheduling and weather confounds, we expect to see more users going out of their way to locations of interest and contributing additional data at target locations and en route.

To better capture users' notification preferences, we designed an improved notification preference survey that provides users with finer-grained control over what notifications they would like to receive (or not); see Figure 5a. Unlike the survey from Study 1, which only allows users to specify individual properties about locations that may be of interest to them (e.g., private tables, windows), this improved survey allows users to specify, for any state of information LES has in its data scaffold, whether they would like to be notified or not. For example, Figure 5a demonstrates how this allows a user to specify that they only want to know about private tables when LES knows that these tables are near windows, which wouldn't have been possible in Study 1. Additionally, users can now select specific locations they would like to be notified about (e.g., notify me about Starbucks, but not Peet's Coffee). With an improved understanding of users' notification preferences, we expect to send fewer notifications at-distance with information that users do not find useful.

⁷All 18 users used LES; 16 completed the post-survey.

To study the extent to which people's interest in the data may influence their decisions to go out of their way and make additional data contributions, we designed an information preference survey that asks users to state how likely they think they are to visit target locations (i.e., Always, Sometimes, Rarely, or Never) when information of interest is presented to them, assuming no scheduling conflicts; see Figure 5b. By collecting both notification preferences and information preferences, we are able to measure the likelihood of users actually going out of their way when notified with information based on their notification preferences, given their reported degree of interest in acting on the data based on their information preferences.

Similar to Study 1, we used user responses to at-distance notifications as an ecological momentary assessment (EMA) [49] to assess whether they found the information presented to them useful and whether they were going out of their way based on the information presented or if they happen to be already going. Beyond EMAs, users completed a post-study survey following the 14-day usage period, which asked them to: (1) reflect on what they found valuable or not valuable about the information presented to them; and (2) why they decided to go out of their way and contribute additional data (or not). We followed the same procedure as Study 1 to code and analyze these qualitative responses.

6.1.3 Measures and Analysis. Similar to Study 1, we measured the number of times users went out of their way and made additional data contributions by examining logged location data to see if users actually went out of their way and made contributions. We then analyzed post-study survey responses to gain a deeper understanding of why they decided to go out of their way and make additional contributions. To evaluate whether people were more likely to go out of their way when presented with information of interest, we used users' information preferences to compute, for each level of interest (i.e., Always, Sometimes, Rarely, or Never), the proportion of times users went to the location when notified.

Similar to Study 1, we collected user reports of perceived disruption. To evaluate the extent to which 4X avoids over-notifying users by only sending notifications at-distance with information of interest, we also compared the number of notifications that 4X sent to the number of notifications that would have been sent had we notified all users within 300 meters of a target location—as the Directed condition did in Study 1—by simulating the Directed condition notification policy with location data collected during the study.

6.2 Results

6.2.1 H3: 4X Gathers Additional Contributions from At-Distance Notifications with Minimal Disruption. 4X yielded 49.09% more data by directing users to target locations of interest, resulting in data contributions made at that target locations and en route to them. Figure 6 shows that users in Study 2 made a total of 82 data contributions, 55 of which are from at-location contributions from eXplore and 27 of which (32.91%) are additional contributions that resulted from users responding to at-distance notifications and making additional contributions at target locations via eXpand (14 times) and en route via eXploit (13 times). These additional contributions helped to increase data fidelity at target locations by adding 25.46% more data above the 55 eXplore contributions, and additionally expanded data coverage at locations en route.

Similar to Study 1, users found 4X to be minimally disruptive; the average reported disruption is 1.88—a number between Never Disruptive and Rarely Disruptive—on a 5-point Likert scale (SD = 0.72). Analyzing the number of notifications sent at-distance, we found that 4X only sent 140 notifications (0.56 per user per day) versus the 1051 at-distance notifications (4.17 per user per day) that would have been sent by a Directed approach. In other words, by sending targeted notifications that are well-matched to people's notification preferences, 4X avoided unnecessarily notifying

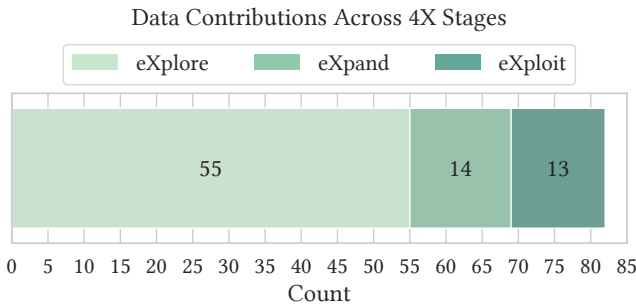


Fig. 6. Total number of contributions in Study 2 by 4X stage. 4X yields 49.09% more data from directing users to locations of interest via eXpand and eXploit over data collected opportunistically via eXplore.

users at-distance who would not be interested in deviating, while still yielding additional data contributions from interested users beyond what can be collected via an Opportunistic approach.

Users generally contributed additional data whenever they went to the target location in response to the data they received (14 out of 17 times for eXpand; 13 out of 14 times for eXploit). Users noted that making additional contributions at the target locations they went to with LES was easy to do and required little effort: *“I picked up cold snacks, and then it asked me what type of cold snack. It was pretty natural to ask a question about something I specifically came for. It was very low-effort to respond, since I already had the knowledge.”* (P16). Similarly, users also felt that it was easy to make contributions en route when LES asked about bike racks and parking lots: *“I went to [the gym] and then was asked about the bike racks. I responded because it was easy for me to respond as it was on my way, and I already had my phone in my hand.”* (P2). These responses show the benefit of using low-effort interaction techniques in 4X systems to collect additional contributions en route to and at target locations that users are drawn to.

Users further highlighted reciprocity as a reason for contributing at eXpand and eXploit locations. For instance, P10 noted: *“I feel like when I get something from others, I would like to give back.”* Beyond providing data back to other data collectors, some users wanted to better know how their contributions were being used by others: *“I wish I knew how my contributions affected others – there were many times I passed [a coffee shop] and was prompted, and I wish I knew how useful my reports were for others, I think it would have motivated me more to maybe step in and check instead of just peer through the windows and guess if I knew someone was relying on the information.”* (P8). Along with providing users with data of interest to motivate additional contributions, 4X systems may also consider informing users about how other data collectors value their contributions as another way to motivate continued use of the system.

6.2.2 H4: Users are More Drawn to Locations of Interest When the Information is Valuable to Them.

As we expected, users were more likely to go out of their way when they found the information presented to them to be valuable. In 16 of the 17 cases in which users went out of their way in response to an eXpand notification, users reported that they would “Always” or “Sometimes” go out of their way. Moreover, Figure 7 shows that when users are presented with the information they find most valuable (i.e., Always), they go out of their way 45.45% of the times versus less than 10% of the time when information is less valuable (i.e., Rarely and Never).

While this result shows that more valuable data has more drawing power, it is worth noting that there are many more cases in which users receive information that is moderately valuable (104 vs. 22), through which six people still went out of their way. From a design perspective, this suggests that while we should generally aim to collect and present information that users value

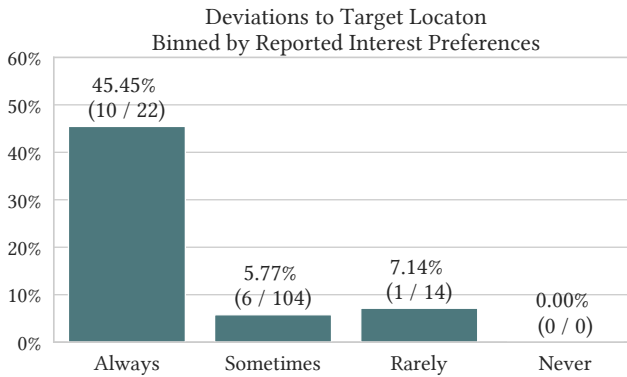


Fig. 7. Users were more likely to go out of their way to locations when presented with information that is more valuable to them.

most, presenting users with moderately valuable data may still be useful and can also lead to additional contributions.

Users reported going to nearby locations of interest that LES notified them about because they generally found the information to be useful. Compared to Study 1, in which 18.42% of eXpand notifications were found to not be useful, only 2.86% of eXpand notifications in Study 2 were found to not be useful. This result suggests that our revised notification preference survey may have better captured users' preferences and thus provided LES with a better user model to use when notifying users about nearby locations. From post-study responses, 9 out of 14 users who went out of their way specifically highlighted that they did so because they found the information they received to be valuable. For example, P16 wrote: *"I was interested in ice cream, as it had been pretty hot and muggy one of the afternoons. I received the notification while at my desk in [building] to visit [another building], and this was interesting enough to stop my work for 15 minutes."* In other words, providing users with data that they generally find useful can be a viable way to draw them to nearby locations of interest outside of their regular routine.

User responses to post-survey questions highlight notification timing as an important factor that influenced their decision to go out of their way and their perceived value of eXpand notifications from LES. P10 noted: *"It depends on timing. In the morning the cafe notification is valuable. In the afternoon after 5 or 6 pm gym availability is more important."* For free food, timing was most often dependent on when the user had last eaten and if they were occupied with work: *"One time I got a notification that there was food. Critically, this notification arrived in the afternoon when I am typically hungry because I do not ever pack enough food for lunch. I was happy to jump up from my desk and find the food because I am also usually less productive in the afternoon."* (P5). These reports suggest that eXpand in 4X may be even more effective in gathering additional contributions and reducing disruption by accounting for such contextual factors when deciding to notify users.

When users did not go out of their way, they primarily cited scheduling conflicts and existing plans as reasons for not going. In the post-survey, 14 out of 16 users noted scheduled meetings to be a reason why they could not go to the location they were notified about. Eight users noted that they were sometimes in a rush (e.g., to catch a train) and thus did not have time to respond. Three users mentioned that they did not want to respond to a notification because they were not interested in the data being collected, suggesting that users' personal interest in the data may influence their willingness to contribute.

Even if users did not go out of their way, some still found value in being provided information about locations of interest through LES as it helped them better plan when they might go to certain locations that they cared about. For example, P14 mentioned how they found the notifications for the gym useful even though they could not go because of work commitments: “*With [the gym], I was glad when I would see open spots, but I was usually at work and could not workout at that time. But at least I would know that the gym is generally empty for when I wanted to go to the gym.*” (P14). This suggests that even when users cannot immediately act upon the data they receive, 4X systems can still provide value to the user as the information could help in planning future trips to locations.

7 DISCUSSION AND FUTURE WORK

Having demonstrated the advantages of 4X over existing opportunistic and directed approaches through LES, we revisit the core ideas behind 4X, discuss how they may generally inform the design of participatory sensing systems, and present directions for future work.

7.1 Keeping Users’ Data Interest at the Foreground

The most important aspect of any 4X system is keeping users’ interests and goals in mind since 4X uses collected data as an incentive for enabling directed sensing to gain additional contributions. Our study results showed that this approach created additional data contribution opportunities (Study 1) and led to increased data contributions both at locations of interest via eXpand and en route via eXploit (Study 2). Unlike the directed approach which makes requests indiscriminately, having user models allowed 4X to selectively notify users only when data of interest is available, which reduced over-notification and limits disruption. In other words, 4X was able to increase data fidelity and coverage in ways that aligned with, and advanced, the interests and goals of users.

Having accurate user models in 4X is essential for selectively notifying users to provide information of interest while avoiding unnecessary disruption during the eXpand stage. In LES, we first conducted needfinding to determine the kinds of data that users are generally interested in about locations and events around campus. We then used a direct elicitation strategy (i.e., a questionnaire) to learn what specific information and events each user was interested in knowing about. While this approach was generally effective for increasing data collection opportunities without increasing disruption, results from our study suggest that including contextual factors such as users’ schedules or how their data interests vary throughout the day may better capture when users are likely to go to a nearby location given some information of interest (as opposed to being interested, but not going). Having a more accurate user model that incorporates such contextual factors would allow a 4X system to reduce disruption from notifications that a user is unlikely to act upon, and may even increase contributions in cases where the model identifies situations where a user is willing to deviate further from their routine in response to information about locations and events that are particularly valuable to them. To build such richer user models, future 4X systems may complement direct elicitation by training machine learning models that predict a user’s likelihood of going out of their way to a target location based on (1) contextual factors such as the user’s schedule and their current activities [12, 18, 21, 27]; (2) their current and future routes [35, 39, 44] and associated cost of diversion to a target location [24, 25]; and (3) the value of the certain, known information about a target location to a user at different times during the day (e.g., line length at a coffee shop in the morning versus late in the afternoon).

While better user models can potentially reduce disruption and increase contributions for some sensing initiatives, the larger issue is that the collected data may not be motivating enough for users to deviate from their routine. For instance, volunteers helping to monitor city infrastructural issues such as potholes and broken streetlights would have no interest in visiting such locations. One approach to collecting such data using 4X is to leverage the eXploit stage to collect mundane

data as a byproduct of a more interesting sensing initiative. As shown in Study 2, LES was able to gather information about bike rack and parking space availability that was unrelated to users' direct interest. This can allow a 4X system that, for example, primarily collects and shares information about events of interest in a city to also collect information about city infrastructural issues that users might encounter en route.

While we do not expect 4X to be useful in all cases, in some cases there may be clever ways to *transform* the collected data so that it becomes incentivizing and useful to data contributors. For example, recent work on Habitsourcing [40] creates immersive interactions and narratives that reference objects in the physical environment to support a user's habit-building practice (e.g., going for a run) while collecting sensing data as a byproduct of their habit building practices. Future Habitsourcing systems may use collected information about locations and events (however mundane) to shape immersive story narratives [46] and interactions that incentivize users to deviate from their routes as part of their habit-building practice, and while doing so, contribute additional data where it is desired. In future work, we are broadly interested in exploring methods such as this for transforming collected data in ways that support interactions and experiences that are valuable to the user, even if the collected data in its raw form is not.

7.2 Using Dynamic Data Collection Processes

4X is a dynamic process for collecting data that uses different data collection mechanisms depending on what data is available to the system and what available information users are interested in. Because 4X leverages the benefits of both opportunistic and directed approaches, it enables the collection of high-fidelity and high-coverage data while being minimally disruptive to users and providing them with data of interest. We show through Study 2 that users made 25% more contributions from eXpand to increase data fidelity and 24% more contributions from eXploit to increase data coverage. By not fixating on a single approach (i.e., opportunistic or directed), 4X is able to *flexibly work with the data and interest available* to achieve a big gain in data contributions while still meeting user needs that would not have been possible with either approach alone.

Dynamic data collection processes such as 4X can be used to scaffold data and motivation concurrently so that larger and larger groups of users can be drawn out of their way to contribute to participatory sensing efforts. We describe below a general strategy for achieving this—which 4X supports—that we call *incentive chaining*. Initially, a small subset of all users, who require the least amount of information to be motivated, are the ones most likely to be drawn out of their routines and make contributions at the target location they were drawn to. Their contributions would then increase data fidelity and motivate other users, who may have needed slightly higher incentives, to now be drawn to the target location and so on. For example, users may pass by an area and opportunistically report an event such as a pop-up concert. This information may then draw in relatively nearby users who are generally interested in any kind of music, where they specify that the concert is for rock music. Then, users who are interested in rock music may be drawn to the location from further away and make additional contributions that may draw in fans of the specific bands playing at the concert from even further away. Thus, incentive chaining and similar strategies that are enabled by dynamic data collection processes such as 4X can expand the set of users who could initially contribute to encompass a much broader set of users as data and motivations scaffold, but only doing so when users can be effectively motivated (e.g., there is now enough data for the user to be drawn in). In large-scale deployments, we expect that a small amount of additional contributions made by a few initial users drawn to locations of interest could lead to a snowball effect of significantly more opportunities created. Studying these incentive chaining effects at scale are a good avenue for future empirical evaluations of 4X-like dynamic data collection processes.

To allow for incentive chaining-like effects to occur, all sensing initiatives need ways to collect the initial pieces of data so that the data collection protocol can dynamically transition from stage to stage. However, some sensing initiatives may find it challenging to obtain these initial opportunistic contributions if, for example, users' normal routines do not coincide with tracked locations. While we used LES to collect all pieces of data for our sensing domain, other initiatives may consider using gamification mechanics [5, 54] or even machine sensors [38, 48] to complement apps such as LES so that particular pieces of data contributed across applications can together support a broader sensing initiative. In other words, the stages of 4X can be enacted across sensing applications, whereby some applications (and their associated mechanisms) make contributions opportunistically, which then empowers other applications to use eXpand and eXploit to further build data fidelity and data coverage. In this way, dynamic data collection processes can easily be extended to integrate multiple forms of participatory sensing while still reaping the benefits of transitioning between data collection stages when users' interests and the collected data align.

7.3 Enabling Community-Based Data Collection

4X systems can be thought of as community-based data collection processes where data contributions from certain community members are of value to other community members, who then become motivated to make further contributions that benefit others in the community and so on. In Study 2, we found that collected data from some users successfully influenced others to deviate from their routine and make additional contributions. In this way, 4X creates cross-community interactions between data collectors since some contributions are being used to inspire contributions from others in the data collection community.

Seeing 4X as a community-wide data collection approach allows us to consider new opportunities for how we may design the data collection process at the community-level rather than only at the individual-level. For instance, in LES we rearranged information scaffolds based on the overall collected user preferences so that data of greatest interest to the community at each location would be collected first. Similarly, we can consider rearranging the information scaffolds based on the needs of smaller sub-communities at different times during the day. For example, those interested in coffee shops may want to know about how long the line at certain shops are during their morning commute, but be more interested in seating availability later in the day. In other words, by explicitly recognizing the needs of different sub-communities within the sensing initiative and tailoring the data collection to their changing needs, 4X is able to find and create more situations where users would be willing to deviate from their routines since the collected data would better align with the sub-community's data needs and interests at different times during the day.

Building 4X systems that continually consider the needs of different sub-communities during data collection requires crafting and tuning more sophisticated notification selection criteria and question selection criteria that can dynamically adjust to the needs of the community based on the information that is available and the people available who may be interested and willing to go out of their way. Manually tuning these selection criteria can become challenging and ineffective when community members have diverse data needs, and when decisions about who to ping now affect opportunities to meet others' needs later. Instead, we are interested in future work that create *Adaptive 4X* systems that can automatically reason and adapt their notification policies based on the value of soliciting certain contributions for the community and of the value of the collected data to people in the community. To best collect data valued by the community without being overly disruptive to any individual, such systems may use decision theory to make decisions about whether to request a data contribution now, or to wait for better opportunities when users may be closer to other locations of interest to them where additional data contributions would be particularly valuable [32]. To best leverage data collectors' efforts and support users' needs across

the community, such systems may adaptively notify more or less users at different times so as to best use the community's data collection efforts to meet the changing data needs of the community (e.g., notify more people who may contribute about coffee shops in the morning). We expect these systems to make decisions that better align with the varying interests of users, which allows for user needs to be better supported while also advancing data collection goals.

8 CONCLUSION

This paper presents 4X: a hybrid, multi-stage approach for participatory sensing that collects low-effort, low-fidelity data opportunistically, and then uses these data to direct users to locations of interest to make additional contributions that build data fidelity and coverage. From two user studies, we found that 4X enables participatory sensing systems to better meet the users' needs for minimal disruption and personally relevant information by only asking them to make contributions at nearby locations that lie outside of their daily routine when certain data of interest is available there, while also better meeting the system's goals for data collection by collecting more data than purely opportunistic data collection approaches would.

Our work provides participatory sensing system designers with insights on designing and implementing multi-stage data collection mechanisms such as 4X into their applications. These insights are important for moving beyond systems that rely on a single data collection process towards systems that draw on multiple sources of motivation for data, and that can better meet data collection goals by using collected data to induce further data contributions. We envision that future participatory sensing systems that build on 4X may continually and intelligently adapt their data collection policies, so that the efforts of a data collection community can best be used to support the needs of data contributors while meeting the data collection goals of the system.

ACKNOWLEDGMENTS

We thank all of our participants for testing our application and providing valuable feedback. We also thank the members of the Design, Technology, and Research program, the Delta Lab, and the CollabLab for helpful discussions. Funding for this research was provided by the National Science Foundation under Grant No. IIS-1618096, and by an Undergraduate Research Grant from Northwestern University.

REFERENCES

- [1] 2019. Building a map for everyone. <https://www.blog.google/products/maps/building-map-everyone/>
- [2] 2019. Field Agent. <https://app.fieldagent.net>
- [3] 2019. Foursquare. <https://foursquare.com/>
- [4] 2019. Gigwalk. <http://www.gigwalk.com/>
- [5] 2019. Ingress Prime. <https://ingress.com/en/>
- [6] 2019. Popular times, wait times, and visit duration. <https://support.google.com/business/answer/6263531?hl=en>
- [7] 2019. SeeClickFix. <https://en.seeclickfix.com/>
- [8] Karl Aberer, Saket Sathé, Dipanjan Chakraborty, Alcherio Martinoli, Guillermo Barrenetxea, Boi Faltings, and Lothar Thiele. 2010. OpenSense: Open Community Driven Sensing of Environment. In *Proceedings of the ACM SIGSPATIAL International Workshop on GeoStreaming (IWGS '10)*. ACM, New York, NY, USA, 39–42. <https://doi.org/10.1145/1878500.1878509>
- [9] Elena Agapie, Jaime Teevan, and Andrés Monroy-Hernández. 2015. Crowdsourcing in the field: A case study using local crowds for event reporting. In *Third AAAI Conference on Human Computation and Crowdsourcing*. <https://www.aaai.org/ocs/index.php/HCOMP/HCOMP15/paper/view/11595>
- [10] Paul Aoki, Allison Woodruff, Baladitya Yellapragada, and Wesley Willett. 2017. Environmental Protection and Agency: Motivations, Capacity, and Goals in Participatory Sensing. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM, New York, NY, USA, 3138–3150. <https://doi.org/10.1145/3025453.3025667>
- [11] Paul M. Aoki, R. J. Honicky, Alan Mainwaring, Chris Myers, Eric Paulos, Sushmita Subramanian, and Allison Woodruff. 2009. A Vehicle for Research: Using Street Sweepers to Explore the Landscape of Environmental Community Action.

- In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '09)*. ACM, New York, NY, USA, 375–384. <https://doi.org/10.1145/1518701.1518762>
- [12] Nikola Banovic, Tofi Buzali, Fanny Chevalier, Jennifer Mankoff, and Anind K. Dey. 2016. Modeling and Understanding Human Routine Behavior. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM, New York, NY, USA, 248–260. <https://doi.org/10.1145/2858036.2858557> event-place: San Jose, California, USA.
 - [13] Jeffrey A. Burke, Deborah Estrin, Mark Hansen, Andrew Parker, Nithya Ramanathan, Sasank Reddy, and Mani B. Srivastava. 2006. Participatory sensing. In *Workshop on World-Sensor-Web (WSW '06): Mobile Device Centric Sensor Networks and Applications*. 117–134. <https://escholarship.org/uc/item/19h777qd>
 - [14] Yung-Ju Chang, Gaurav Paruthi, Hsin-Ying Wu, Hsin-Yu Lin, and Mark W. Newman. 2017. An investigation of using mobile and situated crowdsourcing to collect annotated travel activity data in real-world settings. *International Journal of Human-Computer Studies* 102 (2017), 81 – 102. <https://doi.org/10.1016/j.ijhcs.2016.11.001> Special Issue on Mobile and Situated Crowdsourcing.
 - [15] Dan Cosley, Dan Frankowski, Loren Terveen, and John Riedl. 2007. SuggestBot: Using Intelligent Task Routing to Help People Find Work in Wikipedia. In *Proceedings of the 12th International Conference on Intelligent User Interfaces (IUI '07)*. ACM, New York, NY, USA, 32–41. <https://doi.org/10.1145/1216295.1216309>
 - [16] David Dearman, Timothy Sohn, and Khai N. Truong. 2011. Opportunities Exist: Continuous Discovery of Places to Perform Activities. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11)*. ACM, New York, NY, USA, 2429–2438. <https://doi.org/10.1145/1978942.1979297>
 - [17] Janis L. Dickinson, Benjamin Zuckerberg, and David N. Bonter. 2010. Citizen Science as an Ecological Research Tool: Challenges and Benefits. *Annual Review of Ecology, Evolution, and Systematics* 41, 1 (2010), 149–172. <https://doi.org/10.1146/annurev-ecolsys-102209-144636>
 - [18] Afsaneh Doryab, Victoria Bellotti, Alaaeddine Yousfi, Shuobi Wu, John M. Carroll, and Anind K. Dey. 2017. If It's Convenient: Leveraging Context in Peer-to-Peer Variable Service Transaction Recommendations. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 1, 3, Article 48 (Sep 2017), 28 pages. <https://doi.org/10.1145/3130913>
 - [19] Shane B. Eisenman, Emiliano Miluzzo, Nicholas D. Lane, Ronald A. Peterson, Gahng-Seop Ahn, and Andrew T. Campbell. 2010. BikeNet: A Mobile Sensing System for Cyclist Experience Mapping. *ACM Trans. Sen. Netw.* 6, 1, Article 6 (Jan 2010), 39 pages. <https://doi.org/10.1145/1653760.1653766>
 - [20] Denzil Ferreira, Vassilis Kostakos, and Anind K. Dey. 2015. AWARE: Mobile Context Instrumentation Framework. *Frontiers in ICT* 2 (2015), 6. <https://doi.org/10.3389/fict.2015.00006>
 - [21] James Fogarty, Scott E. Hudson, Christopher G. Atkeson, Daniel Avrahami, Jodi Forlizzi, Sara Kiesler, Johnny C. Lee, and Jie Yang. 2005. Predicting Human Interruptibility with Sensors. *ACM Trans. Comput.-Hum. Interact.* 12, 1 (March 2005), 119–146. <https://doi.org/10.1145/1057237.1057243>
 - [22] Raghu K. Ganti, Nam Pham, Hossein Ahmadi, Saurabh Nangia, and Tarek F. Abdelzaher. 2010. GreenGPS: A Participatory Sensing Fuel-efficient Maps Application. In *Proceedings of the 8th International Conference on Mobile Systems, Applications, and Services (MobiSys '10)*. ACM, New York, NY, USA, 151–164. <https://doi.org/10.1145/1814433.1814450>
 - [23] Raghu K. Ganti, Fan Ye, and Hui Lei. 2011. Mobile crowdsensing: current state and future challenges. *IEEE Communications Magazine* 49, 11 (November 2011), 32–39. <https://doi.org/10.1109/MCOM.2011.6069707>
 - [24] Eric Horvitz, Paul Koch, and Muru Subramani. 2007. Mobile Opportunistic Planning: Methods and Models. In *Proceedings of the 11th International Conference on User Modeling (UM '07)*. Springer-Verlag, Berlin, Heidelberg, 228–237. https://doi.org/10.1007/978-3-540-73078-1_26 event-place: Corfu, Greece.
 - [25] Eric Horvitz and John Krumm. 2012. Some Help on the Way: Opportunistic Routing Under Uncertainty. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing (UbiComp '12)*. ACM, New York, NY, USA, 371–380. <https://doi.org/10.1145/2370216.2370273> event-place: Pittsburgh, Pennsylvania.
 - [26] Gary Hsieh, Ian Li, Anind Dey, Jodi Forlizzi, and Scott E. Hudson. 2008. Using Visualizations to Increase Compliance in Experience Sampling. In *Proceedings of the 10th International Conference on Ubiquitous Computing (UbiComp '08)*. ACM, New York, NY, USA, 164–167. <https://doi.org/10.1145/1409635.1409657>
 - [27] Kazushi Ikeda and Keiichiro Hoashi. 2017. Crowdsourcing GO: Effect of Worker Situation on Mobile Crowdsourcing Performance. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM, New York, NY, USA, 1142–1153. <https://doi.org/10.1145/3025453.3025917>
 - [28] Thivya Kandappu, Nikita Jaiman, Randy Tandriansyah, Archan Misra, Shih-Fen Cheng, Cen Chen, Hoong Chuin Lau, Deepthi Chander, and Koustuv Dasgupta. 2016. TASKer: Behavioral Insights via Campus-based Experimental Mobile Crowd-sourcing. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '16)*. ACM, New York, NY, USA, 392–402. <https://doi.org/10.1145/2971648.2971690>
 - [29] Thivya Kandappu, Archan Misra, Shih-Fen Cheng, Nikita Jaiman, Randy Tandriansyah, Cen Chen, Hoong Chuin Lau, Deepthi Chander, and Koustuv Dasgupta. 2016. Campus-Scale Mobile Crowd-Tasking: Deployment and Behavioral Insights. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing (CSCW '16)*. ACM, New York, NY, USA, 800–812. <https://doi.org/10.1145/2818048.2819995>

- [30] Salil S. Kanhere. 2011. Participatory Sensing: Crowdsourcing Data from Mobile Smartphones in Urban Spaces. In *2011 IEEE 12th International Conference on Mobile Data Management*, Vol. 2. 3–6. <https://doi.org/10.1109/MDM.2011.16>
- [31] Sunyoung Kim, Jennifer Mankoff, and Eric Paulos. 2013. Sensr: Evaluating a Flexible Framework for Authoring Mobile Data-collection Tools for Citizen Science. In *Proceedings of the 2013 Conference on Computer Supported Cooperative Work (CSCW '13)*. ACM, New York, NY, USA, 1453–1462. <https://doi.org/10.1145/2441776.2441940>
- [32] Yongsung Kim, Darren Gergle, and Haoqi Zhang. 2018. Hit-or-Wait: Coordinating Opportunistic Low-effort Contributions to Achieve Global Outcomes in On-the-go Crowdsourcing. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, USA, Article 96, 12 pages. <https://doi.org/10.1145/3173574.3173670>
- [33] Yongsung Kim, Emily Harburg, Shana Azria, Aaron David Shaw, Elizabeth M Gerber, Darren Robert Gergle, and Haoqi Zhang. 2016. Studying the Effects of Task Notification Policies on Participation and Outcomes in On-the-go Crowdsourcing. In *Proceedings, The Fourth AAAI Conference on Human Computation and Crowdsourcing (HCOMP 2016)*. AAAI. <https://www.aaai.org/ocs/index.php/HCOMP/HCOMP16/paper/view/14029>
- [34] Stephen F. King and Paul Brown. 2007. Fix My Street or else: Using the Internet to Voice Local Public Service Concerns. In *Proceedings of the 1st International Conference on Theory and Practice of Electronic Governance (ICEGOV '07)*. ACM, New York, NY, USA, 72–80. <https://doi.org/10.1145/1328057.1328076>
- [35] John Krumm and Eric Horvitz. 2006. Predestination: Inferring Destinations from Partial Trajectories. In *Eighth International Conference on Ubiquitous Computing (UbiComp 2006)*. Springer. <https://www.microsoft.com/en-us/research/publication/predestination-inferring-destinations-from-partial-trajectories/> bibtex*[edition=Eighth International Conference on Ubiquitous Computing (UbiComp 2006)].
- [36] Nicholas D. Lane, Shane B. Eisenman, Mirco Musolesi, Emiliano Miluzzo, and Andrew T. Campbell. 2008. Urban Sensing Systems: Opportunistic or Participatory?. In *Proceedings of the 9th Workshop on Mobile Computing Systems and Applications (HotMobile '08)*. ACM, New York, NY, USA, 11–16. <https://doi.org/10.1145/1411759.1411763>
- [37] Nicholas D. Lane, Emiliano Miluzzo, Hong Lu, Daniel Peebles, Tanzeem Choudhury, and Andrew T. Campbell. 2010. A survey of mobile phone sensing. *IEEE Communications Magazine* 48, 9 (Sep. 2010), 140–150. <https://doi.org/10.1109/MCOM.2010.5560598>
- [38] Gierad Laput, Walter S. Lasecki, Jason Wiese, Robert Xiao, Jeffrey P. Bigham, and Chris Harrison. 2015. Zensors: Adaptive, Rapidly Deployable, Human-Intelligent Sensor Feeds. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. ACM, New York, NY, USA, 1935–1944. <https://doi.org/10.1145/2702123.2702416>
- [39] Lin Liao, Donald J. Patterson, Dieter Fox, and Henry Kautz. 2007. Learning and inferring transportation routines. *Artificial Intelligence* 171, 5 (April 2007), 311–331. <https://doi.org/10.1016/j.artint.2007.01.006>
- [40] Katherine Lin, Henry Spindell, Scott Cambo, Yongsung Kim, and Haoqi Zhang. 2016. Habitsourcing: Sensing the Environment Through Immersive, Habit-Building Experiences. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology (UIST '16)*. ACM, New York, NY, USA, 639–650. <https://doi.org/10.1145/2984511.2984533>
- [41] Thomas Ludwig, Julian Dax, Volkmar Pipek, and Dave Randall. 2016. Work or Leisure? Designing a User-centered Approach for Researching Activity "in the Wild". *Personal Ubiquitous Comput.* 20, 4 (Aug 2016), 487–515. <https://doi.org/10.1007/s00779-016-0935-7>
- [42] Jeffrey Nichols and Jeon-Hyung Kang. 2012. Asking Questions of Targeted Strangers on Social Networks. In *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work (CSCW '12)*. ACM, New York, NY, USA, 999–1002. <https://doi.org/10.1145/2145204.2145352>
- [43] Sangkeun Park, Sujin Kwon, and Uichin Lee. 2018. CampusWatch: Exploring Communitysourced Patrolling with Pervasive Mobile Technology. *Proceedings of the ACM on Human-Computer Interaction* 2, CSCW, Article 134 (Nov 2018), 25 pages. <https://doi.org/10.1145/3274403>
- [44] Donald J. Patterson, Lin Liao, Dieter Fox, and Henry Kautz. 2003. Inferring High-Level Behavior from Low-Level Sensors. In *UbiComp 2003: Ubiquitous Computing*, Gerhard Goos, Juris Hartmanis, Jan van Leeuwen, Anind K. Dey, Albrecht Schmidt, and Joseph F. McCarthy (Eds.). Vol. 2864. Springer Berlin Heidelberg, Berlin, Heidelberg, 73–89. https://doi.org/10.1007/978-3-540-39653-6_6
- [45] Reid Priedhorsky, Mikhail Masli, and Loren Terveen. 2010. Eliciting and Focusing Geographic Volunteer Work. In *Proceedings of the 2010 ACM Conference on Computer Supported Cooperative Work (CSCW '10)*. ACM, New York, NY, USA, 61–70. <https://doi.org/10.1145/1718918.1718931>
- [46] Mark Owen Riedl and Vadim Bulitko. 2013. Interactive Narrative: An Intelligent Systems Approach. *AI Magazine* 34, 1 (2013), 67–67. <https://doi.org/10.1609/aimag.v34i1.2449>
- [47] Adam Sadilek, John Krumm, and Eric Horvitz. 2013. Crowdphysics: Planned and Opportunistic Crowdsourcing for Physical Tasks. In *Seventh International AAAI Conference on Weblogs and Social Media*. <https://www.aaai.org/ocs/index.php/ICWSM/ICWSM13/paper/view/6121>

- [48] Manaswi Saha, Michael Saugstad, Hanuma Teja Maddali, Aileen Zeng, Ryan Holland, Steven Bower, Aditya Dash, Sage Chen, Anthony Li, Kotaro Hara, and Jon Froehlich. 2019. Project Sidewalk: A Web-based Crowdsourcing Tool for Collecting Sidewalk Accessibility Data At Scale. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. ACM, New York, NY, USA, Article 62, 14 pages. <https://doi.org/10.1145/3290605.3300292>
- [49] Saul Shiffman, Arthur A. Stone, and Michael R. Hufford. 2008. Ecological Momentary Assessment. *Annual Review of Clinical Psychology* 4, 1 (2008), 1–32. <https://doi.org/10.1146/annurev.clinpsy.3.022806.091415> PMID: 18509902.
- [50] Brian L. Sullivan, Christopher L. Wood, Marshall J. Iliff, Rick E. Bonney, Daniel Fink, and Steve Kelling. 2009. eBird: A citizen-based bird observation network in the biological sciences. *Biological Conservation* 142, 10 (2009), 2282 – 2292. <https://doi.org/10.1016/j.biocon.2009.05.006>
- [51] H. Tangmunarunkit, C. K. Hsieh, B. Longstaff, S. Nolen, J. Jenkins, C. Ketcham, J. Selsky, F. Alquaddoomi, D. George, J. Kang, Z. Khalapyan, J. Ooms, N. Ramanathan, and D. Estrin. 2015. Ohmage: A General and Extensible End-to-End Participatory Sensing Platform. *ACM Trans. Intell. Syst. Technol.* 6, 3, Article 38 (Apr 2015), 21 pages. <https://doi.org/10.1145/2717318>
- [52] Anthony Tomic, John Zimmerman, Aaron Steinfeld, and Yun Huang. 2014. Motivating Contribution in a Participatory Sensing System via Quid-pro-quo. In *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW '14)*. ACM, New York, NY, USA, 979–988. <https://doi.org/10.1145/2531602.2531705>
- [53] Khai N. Truong, Thariq Shhipar, and Daniel J. Wigdor. 2014. Slide to X: Unlocking the Potential of Smartphone Unlocking. In *Proceedings of the 32nd Annual ACM Conference on Human Factors in Computing Systems (CHI '14)*. ACM, New York, NY, USA, 3635–3644. <https://doi.org/10.1145/2556288.2557044>
- [54] Kathleen Tuite, Noah Snaveley, Dun-yu Hsiao, Nadine Tabing, and Zoran Popovic. 2011. PhotoCity: Training Experts at Large-scale Image Acquisition Through a Competitive Game. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11)*. ACM, New York, NY, USA, 1383–1392. <https://doi.org/10.1145/1978942.1979146>
- [55] Heli Vääätäjä and Paul Egglestone. 2012. Briefing News Reporting with Mobile Assignments: Perceptions, Needs and Challenges. In *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work (CSCW '12)*. ACM, New York, NY, USA, 485–494. <https://doi.org/10.1145/2145204.2145280>
- [56] Rajan Vaish, Keith Wyngarden, Jingshu Chen, Brandon Cheung, and Michael S. Bernstein. 2014. Twitch Crowdsourcing: Crowd Contributions in Short Bursts of Time. In *Proceedings of the 32nd Annual ACM Conference on Human Factors in Computing Systems (CHI '14)*. ACM, New York, NY, USA, 3645–3654. <https://doi.org/10.1145/2556288.2556996>
- [57] Jiangtao Wang, Yasha Wang, Daqing Zhang, Feng Wang, Yuanduo He, and Liantao Ma. 2017. PSAllocator: Multi-Task Allocation for Participatory Sensing with Sensing Capability Constraints. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW '17)*. ACM, New York, NY, USA, 1139–1151. <https://doi.org/10.1145/2998181.2998193>
- [58] John Zimmerman, Anthony Tomic, Charles Garrod, Daisy Yoo, Chaya Hiruncharoenvate, Rafae Aziz, Nikhil Ravi Thiruvengadam, Yun Huang, and Aaron Steinfeld. 2011. Field Trial of Tiramisu: Crowd-sourcing Bus Arrival Times to Spur Co-design. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11)*. ACM, New York, NY, USA, 1677–1686. <https://doi.org/10.1145/1978942.1979187>

Received April 2019; revised June 2019; accepted August 2019