

The Potential of Computer-Assisted Direct Observation Apps

<http://dx.doi.org/10.3991/ijim.v9i1.4205>

D. Wessel

Leibniz-Institut für Wissensmedien, Tübingen, Germany

Abstract—Direct behavior observation, i.e., without first creating a video recording, is a challenging, one-shot task. The behavior has to be coded accurately during the situation itself. Mobile devices can assist direct observation, and there already are applications available for these purposes. However, the mobile revolution has led to new developments in devices, infrastructure, and market penetration that have opened up new, yet untapped, possibilities. In this article, expanded activity theory is used to highlight the unused potential of computer assisted direct observation (CADO) apps. If this potential is realized, it can provide observation with the same advantages online questionnaires and sites like Mechanical Turk have provided for surveys and Internet experiments.

Index Terms—computerized observation, data collection, observation systems, mobile devices, computer-assisted direct observation programs.

I. INTRODUCTION

Observational studies are used in many scientific disciplines, among others, applied behavior analysis, child development, animal behavior [1], and visitor studies [2]. They provide important information about actual behavior ([3]; [4]; [2]) with a high level of detail [2] and without being dependent on language [4]. They also allow for more contextual information in the field [5] and are highly sensitive regarding detailed events [6]. If they are done unobtrusively, they avoid reactivity problems like the guinea pig effect, role selection, or measurement as a change agent (see, e.g., [7]; [4]). Observations can also be used alone or in concert with other methods like surveys or interviews.

However, observations are challenging. While patterns in behavior can be identified [8], the researcher has to choose, usually based on theory, on which behaviors and behavioral patterns to focus. In quantitatively oriented observations, these behaviors are tracked and quantified according to frequency and time [9]. For achieving reliable and valid data, observations place high demands on the observers' attention, accuracy, and speed to notice and accurately record these events. Biases like human observer or investigator errors [4] threaten the internal and external validity of the observations. Observer agreement, the "sine qua non of observational research" [10] can usually only be determined after the observation, and only if multiple observers or observations are available (intraobserver and interobserver agreement, cf. [1]). It is not an easy task to conduct valid observations under economic and practical constraints (e.g., feasibility, operating ease, and cost-benefit ratio; see, e.g., [4]).

One way to deal with these demands is to use video recording. However, whereas a video recording of the ob-

servations can reduce time-pressure and improve accuracy and reliability, there are many situations that prevent video from being used. Privacy norms and expectations are an increasingly important issue. Depending on country and setting, laws might prohibit unobtrusive, or even obtrusive, recording. Ethical concerns make covert video recording questionable ([11]; [2]), whereas informing the subjects prior to the observation incurs reactivity issues. There are also practical concerns, like the difficulty in covering a whole area with cameras [2], or getting high quality audio/video data [11]. A fixed camera with its 2D view is also not as flexible as a human observer [12]. These issues are more prominent in naturalistic observations of human subjects, which frequently are of particular interest.

Manually creating a record of the observations during the situation itself is often the only feasible way. However, with no video recording to refer to, direct observation is a challenging, "one-shot" task. Similar to the "decisive moment" in photography [13], a missed event cannot be repeated. Thus there is a strong need for methods to conduct observations in the situation itself that are as accurate, reliable, valid and efficient as possible. Frequently, pencil and paper are used for direct observations. But as [2] pointed out, despite the ostensibly low cost and ease of use, paper and pencil have several disadvantages: It is difficult to log accurate times for overlapping actions, writing on clipboards is highly obtrusive, and digitizing the data requires additional effort.

Handhelds have become common for specialized commercial and scientific purposes in the last two decades ([14]; [15]). Many advantages of digital technology for observations are acknowledged. For example, [2] and others point to the potential of digital technology for observations, esp. handheld computer technology or "computer-assisted direct observation programs" [15]. Compared to pencil and paper, these digital devices are extremely useful for observation. By using virtual buttons and automatic logging to code behavior and behavior patterns, they allow for more accurate recording with separate times for concurrent behaviors ([1]; [2]), are easy to learn ([16]; [2]), eliminate the need to digitize/transcribe the data [2] and thus help to avoid possible transcription errors [3]. They are also harder for the observed subject to notice [2], allow the observer to code more categories of behaviors ([1]; [16]), can assist in the observation (e.g., via questions, [16]) and provide a means of checking whether data was entered completely and at the correct time [3]. Especially the elimination of transcription costs can make mobile devices cost-effective, despite their higher initial costs ([16]; [2]). Main cost factors are devices and software licenses for the observation app, occasional replacement costs [3], and training

costs [16]. There are even developments that use voice recognition technology to automatically translate spoken observations into text [17].

The utility of a tool depends on the specific goal of the observation, yet compared to video recording, mobile media allow for more flexible positioning of the observer, e.g., in tracking subjects over larger areas and independent from a fixed camera angle [2]. Digital devices provide the observer with more contextual information when doing the coding [12]. Privacy, legal, and ethical issues are also avoided or reduced through their use.

The theoretical advantages of digital devices do indeed seem to turn into practical ones [3]. The technical issues of using mobile devices seem to be under control and are well understood. Guidelines for user interface are widely available (e.g., Apple UI Guidelines, see [18]). Important criteria like accuracy, speed, error correction, and mechanical and environmental factors (battery life, screen visibility, resistance to adverse weather conditions) are identified ([19]; [3]; [17]).

However, despite their positive aspects, the full potential of mobile devices has barely been tapped. While the systems have become more powerful and user-friendly as [14] predicted nearly 20 years ago, developments in devices, infrastructure, and market penetration have also opened up new, yet untapped, possibilities. The change has not simply been a quantitative change in terms of device power, but a qualitative change as well. The mobile revolution has vastly improved technology, added new features, and made the devices ubiquitously available. From specialized, organizer-like devices, mobile media have become all-around mobile computers with touchscreen-based user interfaces. Other powerful hardware features allow for powerful new uses. The potential of such features can assist direct observers in using that one opportunity to capture behavior and improve observations in general.

In the present paper, the focus is on this new potential in the hope of stimulating further development of computer-assisted direct-observation apps (CADO apps). CADO apps are defined as apps running on commercial tablets and smartphones to assist users in scientific observations of other people, animals, or processes.

To identify potential advantages, we have adapted the expanded activity theory for mobile learning ([20]; [21]) to CADO apps. Whereas observation is different from mobile learning, this theory sheds light on the use of mobile technology to facilitate an outcome, here to ensure or improve a psychological process like learning. Thus, expanded activity theory seems to be a suitable basis for identifying the potential of CADO apps. After a short overview of the theory, the factors of this theory will guide the main part of the paper.

II. EXPANDED ACTIVITY THEORY

The expanded activity theory for mobile learning ([20]; [21]) was designed to structure and analyze mobile learning and has also been used for a state of the art literature review [22]. Given a different focus — in this case, mobile learning vs. conducting observations — some adaptation is needed. But since activity theory has an entire “activity” as a unit of analysis [23], instead of a specific tool or purpose, the theory can be adapted for computer assisted direct observation apps. Activity theory allows “captur-

ing the myriad possible interactions that learners may engage in as they roam around their respective environments” [20], and likewise, it can be used to look for unused potential. Thus, activity theory is suited to show “points of attack to improve tool-mediated activity” [23].

Activity theory for mobile learning consists of six interconnected factors: tools, subject, objective, control, context, and communication. Additionally, there are two perspectives, or layers, from which to look at these six factors: the semiotic layer (“the learner’s object-oriented actions [...] are mediated by cultural tools and signs”, [21]) and the technological layer. For a description of six factors and two layers in context of CADO apps, see Table 1.

TABLE I.
ACTIVITY THEORY APPLIED TO MOBILE OBSERVATION

	Technological Layer	Semiotic Layer
Tool	e.g., iPod touch, iPhone	observation scheme
Subject	technology user	observer
Objective	entering valid observation data	learning about behavior
Control	usability	social rules of the observation setting; privacy regulations
Context	physical environment, incl. visibility of behavior	observer/research community
Communication	communication functions, e.g., via chat	exchange of observation information and skills

The overall desired consequence of the activity (the changed object) is the improvement of observation quality or its maintenance at a very high level.

Whereas the semiotic and technological layers can be viewed separately, both layers closely and dynamically interact with each other [21]. Thus, it makes sense to look at them together, to examine the interaction between the user and the applied technology and the consequences of this interaction for improving observation.

III. LOOKING AT THE DIFFERENT FACTORS FOR THEIR USE IN COMPUTER-ASSISTED DIRECT OBSERVATION

The sections on each factor start with a short description and are followed by relating the factor to ways to improve observation practice.

A. Tool

Mobile devices usually share a number of attributes: They are portable individual devices that can react to the current context (context sensitivity) and allow face-to-face and computer-mediated sharing of information (cf. [24]). Whereas users can show the screen to others, they can also easily use it in private, hiding it from view (more in “Subject” and “Objective”). Electronic communication includes, for example, Bluetooth, WLAN, or Cellular data (more in “Communication”), while context sensitivity includes sensors like, e.g., GPS, accelerometer, compass, microphone, camera (more in “Context”). With these sensors, the device can determine location, time, movement, and more. Hardware sensors (e.g., Pasco, see [25]) can also be connected to smartphones to gather additional data (see “Context”).

Additionally, whereas many mobile devices are marketed as smartphones or music players/entertainment devices (e.g., iPod touch), they actually are fully-fledged and powerful computers with a lot of processing power (see “Control”).

The consequences of most of these features are discussed in the following subsections. However, it is worth pointing out that most of today’s mobile devices come with a virtual interface that can display any visual information and can react to many different kinds of input (simple touch to multitouch gestures). Thus, the virtual interface allows for a high degree of flexibility in displaying information and receiving user input. It allows for natural finger interaction and can be adapted to situational requirements. On a virtual interface, almost anything can be displayed and used as virtual button.

B. Subject

One of the main changes due to the mobile revolution is the ubiquity of mobile devices. They are widely used in most of the world. This allows for a large user base of potential observers if the observers’ own devices are used, akin to the “bring your own device” (BYOD) movement in education. Using the observers’ own devices can also save researchers from having to allocate resources for the hardware (see [26]). Likewise, the development of apps for mobile devices has become popular. Powerful integrated development environments like XCode and widely available online instructions and communities (e.g., on Stackoverflow or MOOCs about programming) make it possible for laypeople to develop apps.

Thus, with regard to the subject in the activity, here the observer, the potential advantages of CADO apps include improved personalization, the involvement of lay observers (similar to the use of survey respondents with Mechanical Turk), and keeping the observer engaged.

1) Improved Personalization

Mobile media usually are personal devices. Thus they can be configured to suit the individual user (see [24]). This might seem at odds with reliability, the “sine qua non of observational research” [10]. However, it is not important that the observation tool looks the same for each observer, but that different observers produce identical (and valid) observation data. The interface should be designed well, but adaptations should also be possible.

On the software side, if a virtual interface is user-editable, it allows for changes in the placement and size of buttons, or mirroring the whole layout to change handedness. Given that handedness is an important consideration in handheld vigilance tests [27], taking handedness into account should lead to improved accuracy and in turn improved reliability, validity and efficiency as well. Language settings are another area where flexibility is useful. On a more strategic level, the number of possible observational categories can be adapted to the skill of the observer. Using a basic set of categories for all observers ensures comparability, while additional or more detailed categories can be used as “add-on” for more experienced observers. When it comes to output, personal reminders via prompts and feedback can be given according to the observer’s skill level (as implemented for video observation/coding by [1]).

On the hardware side, given the ubiquity and frequent usage of mobile devices, observers likely are more famil-

iar and comfortable with using their own devices for observations. They do not need to adapt to handling another device.

2) Use of Lay Observers

The ubiquity of mobile devices combined with allowing observers to use their own devices can open up observation to laypeople. Whereas some disciplines have a long tradition of non-professionals as observers (e.g., astronomy or ornithology), other disciplines usually use non-professionals mainly as study subjects to be observed. However, CADO Apps on ubiquitously available devices can do for observations what the Internet has done for surveys: Provide access to data created by people from all over the world. Observers can be recruited via sites like Mechanical Turk (<https://www.mturk.com>), or can be volunteers like students in a school/university class or interested citizen scientists (cf. [28]). This allows for observations in different contexts and of different groups (incl. frequently neglected ones, see, e.g., [6]). Informants (see [4]) can be used who have access to otherwise inaccessible settings.

Recruiting interested citizen scientists or paid lay-observers can thus vastly increase the scope and generalizability of observations. This possibility addresses the frequent concern that some areas of science are focused too much on artificial situations and lack generalizability. Some applications already point to the potential for using them in developing countries (e.g., [16]) and laypeople are already involved in some observations [5]. Note that legal and ethical issues must be taken into account, both in the country where the observations take place and in the country that is responsible for the research.

Using lay people requires that the App (and the specific interface used for the observation) are available on the commercial devices. Training and evaluation are also needed to ensure that only laypeople who are able to use the app and meet the required level of skill can become observers (see “Control”). Video data with volunteers can be used to describe the observable behaviors. As for trust in the data, context-awareness (see “Context”) can be used to ensure that the observations were actually done in the described conditions and in a realistic way. For example, protected log files can be used to check that the data was entered both at the correct time and place and in a realistic time frame (for time implemented by [3]). Interruptions in using the App can be monitored as well, e.g., a manual termination of the app or an incoming call. Being able to correct user errors easily is important, but editing older data should require supervisor access to prevent retrospective entry. These features can lend trust in the quality of the data and expose data fabrication and manipulation. Since the device time can be changed, the app should “call home” via a short message when it is used (see “Connectivity”).

3) Keeping the Observer Engaged

Given the repetitiveness and downtime involved in many observations and the resulting focused attention that is required, some degree of observer engagement is desirable. For example, gamification (e.g., [29]) can be used to make the observation more engaging by, e.g., indicating the amount of codings done in intervals or providing feedback on one’s competence as observer. The goal here is not to turn a scientific observation into a game but to ensure high levels of attention. Thus, these “add-ons” must

never interfere with the observation and must not lead to observers' "gaming the game" (i.e., trying to get a high score at the cost of observation quality; cf. [29]). The advantages of a mobile platform can be used to keep the observer engaged. The mobile platform can run multiple apps, so other apps can be used during downtime. An example of such a downtime might be during the time a subject is in an area that should not be observed (e.g., a café in a museum) and the observer has to wait for the subject to leave that area. However, whereas the time outside the observation app can be easily tracked by logging when the app goes into the background and returns to the foreground, the observer must retain enough attention to notice when the observation must resume (i.e., when the subject should be observed again).

C. Objective

Looking specifically at how CADO apps can support the activity's objective (improved/high validity and information value of observations), the potential includes improvement of the unobtrusiveness of observations, increase of trust by protecting the observed person's privacy (in observations with humans), and provision of additional information about the observation targets.

1) Improvement of Unobtrusiveness of Observation

If observers are discovered, it threatens internal and external validity (see [4]). Reactivity comes into play, which, while predictable in some cases (see [6]), is best avoided. Given the size and ubiquitousness of mobile devices, they do not draw much attention and can easily be hidden (see for example Fig. 1, used in an exploratory study in an art museum). Additionally, dummy screens (e.g., the app displaying an image of a Wikipedia App which is done by some "spy-cam" apps) can be used when the observed person comes too close and can look at the screen.

2) Increase of Trust by Protecting the Observed Person's Privacy

Even without use of video, the observation data might be potentially damaging to the observed person (e.g., observing mistakes made by employees). Given that the app handles the data, the data can be immediately encrypted (also highly relevant for data transmission, see "Communication"). Alternatively, it is possible to prevent displaying the data on the device itself. If an observed person objects to being observed, the data can still be purged, e.g., by using a consecutive observation number, a code, and/or date/time information to identify and delete the relevant data. Accessing aggregated data for on-the-fly analyses (see "control") or using it to generate custom questions (see below) is still possible.

3) Provision of Additional Information about the Observed "Objects"

The computational power combined with digital data affords lots of potential for detecting interesting patterns by using algorithms for pattern detection (see, e.g., [8]). Done during or immediately after the observation it allows for confronting a human subject with his/her idiosyncratic behavior. Recorded behavior can be dynamically replayed or summarized on the device. Replays can include using a map for displaying movement and short texts/icons for observed behaviors. Summaries can include a heat-map of the target's location (see [30], for museum evaluation reports) or a list of the most or least frequent, longest or



Figure 1. An Observation Tool embedded in a Paper Notebook

shortest behaviors. If "prototypical" behavior data is available as a comparison (e.g., already collected and aggregated data), atypical behavior or rare combinations of behaviors can be examined in detail by pointing them out as well.

Think-aloud protocols can be used to determine the targets' explanations of their behavior. If allowed, the explanations of the subjects can be recorded via the microphone of the smartphone and — in a dynamic replay — stored with the respective time codes for later analysis. For animals, further analysis might also be suggested.

To avoid observer bias, computer-assisted self-administered interviews [31] can be used for humans. The observed person answers the questionnaire on the device, similar to a web-survey, but without the need for a network connection. Given that the behavior information is stored on the device, the data can be used to select and frame questions, similarly to adaptive testing.

D. Control

Regarding the control factor of activity theory, a balance must be struck between control by the device and control by the observer. If handled well, CADO apps can improve usability by preventing observation errors. This support can be helpful to compensate for lack of experience among observers (i.e., lay observers). However, when limitations cannot be overcome by the app itself, it might be necessary to go beyond the app. In particular, the focus here is on using visual representations and strong feedback, active error checking and prevention, and using handwritten annotations for rare events.

1) Using Visual Representations and Strong Feedback

Given the visual interface on which almost anything can be displayed, visual representations can be used for tracking. For example, similar to the presentation of observation results in evaluation reports (e.g., [30]), a map can be used for tracking as well. To assist in tracking, the displayed map can also be divided into a chessboard or hexagonal pattern (see Fig. 2). Alternatively, areas of interest defined by the setting itself can be used, e.g., exhibits in a museum, natural objects like water holes in the wild, or play objects in a cage. Larger settings can be broken down into sub-sections accessible via taps or by implementing zooming gestures. Likewise, buttons can represent the actual objects or behaviors. For instance, in a

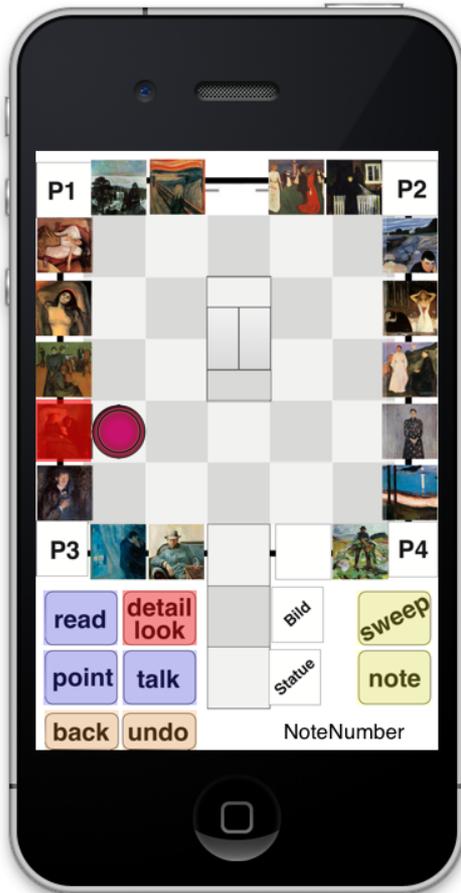


Figure 2. Visual Representation of the Setting and the Objects in an Observation in a working App prototype

gallery, artworks can be represented with small images/icons of these artworks (see Fig. 2). Likewise, photos of specific physical markings can be used to help identifying animals. The camera of the device can be used to create/modify the buttons.

To facilitate observations, buttons can be grouped as mutually exclusive actions. A new action within the same group ends the previous action. Likewise, for one-time actions buttons can “light up” for a few seconds after being pressed and then slowly fade out. Buttons can also stay pressed (highlighted) for events that extend in time/continuing actions (similar to a latching switch). Audio-feedback can be used when a button is pressed, ranging from a simply click sound (mimicking physical buttons) to a sound file denoting the behavior. Earphones prevent disturbing the subject or inducing reactivity.

To use the limited screen real-estate optimally, only currently relevant buttons should be displayed. Also, buttons can modify the behavior of other buttons. For example, during a visitor observation, selecting a new exhibit deactivates the previously selected exhibit, as a visitor usually looks only at one exhibit at a time. However, if a “sweep gaze” button is pressed, multiple exhibits can be selected (when the visitor looks at multiple ones at the same time). Using multitouch would be another way to deal with multiple actions at the same time, if the buttons are close together. “Action windows” (see [16]) can be used to offer more detailed choices for categories of behavior without overcrowding the small display. Depending on the information, either the time of the first button

press or of the detailed button press can be used as start time for the behavior. For example, pressing a “talk” button can change the view to allow for more detailed choices (e.g., how loudly or to whom). The time of the first button press is then used as start time for the behavior, given that the detailed button only further specifies the behavior.

2) Active Error Checking and Prevention

The CADO App can assist observers in controlling for errors in multiple ways, starting with preventing biases of whom to observe in the first place. Humans usually have difficulties making random choices and might prefer to observe certain people and not others. While an “imaginary line” [2] can be used to select, e.g., the visitor to track in a museum, generating random choices on the App might lead to a more objectively random selection.

Completeness checks and logic checks (e.g., questioning whether a combination is actually possible) can be used. For example, incomplete observations can be pointed out. Likewise, mutually exclusive actions can be prevented or signaled, e.g., if a certain position does not allow the observed person to look at a specific object. Auto-correction can also be used. However, it is crucial to leave the control with the user. The observer should never “fight” the automated actions of the instrument.

Going further, pattern analysis can be used to indicate deviations from the norm (i.e., data from previous observations) for further investigation. While there is no “typical” subject in many areas (e.g., for museum visitors, see [32]), some deviations might at least hint at observer errors or at special cases that warrant closer observation. For example, actions that deviate in frequency or duration by, e.g., two standard deviations, might be valid, but it is probably good practice to check for an error. Analyzing the observational data during the observation can assist in checking the coding and identifying of errors while they happen. If the deviations are due to coding errors, looking at the amount of corrections done by the observer might point to fatigue or general lack of attention. Thus, longer pauses can be suggested or enforced by the device. Otherwise, actual outliers can trigger additional actions to find out exactly why this subject differs. For example, the rare behavior can be made a topic of an interview (cf. objective above).

3) Using Handwritten Annotations for Rare Events

In the case of unforeseen events, notes are necessary for later reference. As typing on digital devices is a chore for many users, handwritten notes might be preferable. This can be done on the device itself, as handwriting is possible with special pens for capacity touchscreens. Given the virtual display, a button can display a clear writing surface for quick notes and then store the note with its time code. Another solution is the use of a button that provides a consecutive number with each press. This number is logged in the app as an observation with its own time-code. The observer simply writes the number and the notes on a piece of paper. Rare or unexpected behavior can thus be quickly captured with accurate time information. Combining the handwritten notes with the data afterwards is easy, as the number links them. A digital pen [33] might be useful to more quickly digitize the information. Post-its can be used to allow for handwritten notes if the device is hidden in a book, as they provide easily replaceable writing spaces (see Fig. 1).

E. Context

The context in which behavior is shown, including environmental conditions, is an important factor when it comes to understanding behavior [1]. Especially dynamically changing contexts (see [1]) can be tracked easily with mobile devices.

1) *Observation in Field Settings*

Digital devices might not seem suitable for field settings, due to the dangers to electronics posed by water, dust, and dirt. However, cases for outdoor use exist for many modern smartphones. These cases protect the device while still allowing the user to operate the touchscreen. Protection includes waterproofing and impact resistance, as well as protection against dust, dirt and snow. There are also specially designed gloves that can be used in cold climates and which will work on the display. A capacitive stylus also allows data entry when wearing ski gloves or mittens [3]. However, a capacitive touchscreen will not work when wet (rain, emerged in water). This additional equipment needs to be taken into account when planning the observation budget and determining the observation conditions.

However, other advantages of digital devices make up for these limitations. For example, handhelds can be used with one hand alone if the interface is designed accordingly. The fingers hold the device while the thumb presses the virtual buttons. In practice, this allows users to observe behavior in situations where they need their other hand, e.g., for holding themselves in position. Also, a night view mode can be used for observation in low light conditions. Using a simplified interface with red colors for button outlines might help with observations in the dark, by allowing logging observations and helping the observer adapt to the dark. This feature has already been implemented in some apps, e.g., for reading in the dark, or for stargazing (e.g., [34]).

2) *Capturing the Environment*

Many Smartphones have ways to determine the current location via GPS, but also by using available WiFi or Bluetooth signals. While each system has its advantages and limitations, the accuracy is getting better. There are promising developments especially in the area of museum or tour guides that can be used in aiding observations (for a discussion of the potential of these systems see [32]).

While using video observation might not be possible, the camera can be used to capture the environment in which the observation takes place. Many smartphones have high quality photo/video cameras. Photo or video data attached to an observation can provide useful data about the environment -- even without the observation target in it.

Existing capabilities of smartphones can be augmented with external sensors. Scientific sensors can be connected to smartphones (e.g. from Pasco, see [25]). These sensors provide information about the natural environment (e.g., water quality, light, soil properties, temperature, weather, radiation) or, in cued observations, the physiology of the observed person (e.g., blood pressure, respiration). The area of passive telemetrics can provide other useful sensor information for observing human subjects (see, e.g., [7]). Sensors embedded in the environment, e.g., movement sensors in museums to measure visitor density (see, e.g., [4]), also provide useful context information. Sensor data

greatly reduce the amount of work that needs to be done by the observer and increase accuracy and reliability.

3) *Field Analysis and Adaptation*

In some field settings it might be necessary to make adaptations of the observation scheme on the device itself. The visual interface allows these modifications on the device itself and also creates opportunity to discuss handling or coding issues in the actual environment. Importing photos for maps, button images, etc. or drawing areas of interests would be a requirement. This facilitates pilot testing and modifying the instrument. If there are irregularities or elements are grouped too closely together and have to be treated as a single unit (e.g., exhibit elements that cannot be differentiated by an observer, [2]) these positions/groupings can be changed on the fly. The device can also be used to log more general information. Freedom of movement necessitates having to make a choice about the position from which to observe. It can be difficult to find positions from which a user can observe unobtrusively, so it is useful to be able to highlight places on the map where the observer is hidden, yet has a good line of sight. Additionally, this information allows automatically determining situations where no observations are possible.

Information about useful changes can come from analyses done on the device — given the available processing power of mobile devices. These analyses can, e.g., assist in observing specific behaviors in more detail by indicating times and places where these behaviors are frequent. Changes in population stability over time or in certain areas (see [4]) can also be indicated. Similarly, if data from prior observations exist, this data can be used for comparison to the current observation. While individual behavior is potentially unique and fluctuates over time and situations [7], aggregated data from past and current observations should be similar — at least on some key indicators or “marker behaviors.” Deviation from easy to code and relatively unchanging “marker behaviors” can indicate observation problems and suggest the need to check for error in the situation.

4) *Observation in the Scientific Community*

CADO apps can facilitate sharing of observation interfaces, coding schemes, and data. After creation, the interface can be exported to other devices. Likewise, creation of an interface can be done on an external computer and imported into the app. Defining an interface via tables and/or scripts can allow laypeople to create interfaces. Handling the interface separately from the actual app allows for storing different setups in the app.

If the interface and instructions can be imported/exported, it becomes possible to share coding schemes more easily. This makes it easier to replicate observations, and save time in developing coding schemes. However, coding schemes usually are not eagerly adopted by other researchers [35] and given their close ties to theory it is questionable whether they can be easily adopted (e.g., mentioned by [9]). However, importing an existing coding scheme complete with instructions and examples might be a good start.

The way data is handled for analysis can differ. As digital data is highly flexible, an observation app should provide multiple output formats, regarding the used syntax as well as regarding the file format. A syntax editor with placeholders for times and actions would provide the

necessary flexibility to export the data, for example, in a state-object-location-activity syntax [12] or Theme's notation [8]. Data files should also contain meta data about the observation itself, the used interface, available behaviors, etc.

F. Communication

With WiFi and cellular data, handhelds can connect to other devices and data sources, including the Internet (see [24]). While connectivity has tremendous potential, in many observations it should be an option and not a requirement. Otherwise, even an observation backed up by redundancy depends on a working network connection; something the observer usually has little control over. Improvements include communication with the environment, among observers for training and support, among observers for coordination, and with the lead researcher.

1) Communication with the environment

Connectivity allows the device to gather data from the environment. Besides sensor data (see "Context"), devices in the environment can provide data as well. For example, digital exhibits in a museum can communicate usage data when an observed person uses it to further augment the observation data. From another viewpoint toward the observer, in structured observations the observer might want to influence the situation. Connectivity can be used to trigger specific actions in the environment, e.g., giving food to animals, or (re-)starting a movie for visitors in a museum. This trigger can be given manually or based on specific behaviors that were coded (e.g., position of the subject).

2) Communication among observers for training and support

Training observers is crucial for good observation data. Especially if laypeople are used (cf. subject), training should be directly on the devices of the observers and highly automated. For example, sample videos can be provided online, which the potential lay observer codes with the app. Given that the actions are known, the device can then compare the observers codings with the best practice solution as well as highlight misses and false alarms. If the device and the video are synced, problematic situations can be instantly replayed and ambiguous behaviors discussed.

If human behavior is involved, more practical and realistic training is possible. While letting observers observe each other for training purposes is useful [2], there is usually little feedback. Using one device to give orders via headphones (either randomly from a list of behaviors or via a predetermined script) and another connected device to record these behaviors can provide this feedback (akin to Improv Everywhere's mp3 'experiments', [36]). As the actions are known, feedback can be given immediately or at the end of the observation period. Deviations from the best practice solution can either be indicated immediately or after the observation, e.g., as a list or via interobserver agreement values (see [1]). Care must be taken to select a suitable time lag [16] for the best practice solution. It must be large enough to factor in reaction times of both volunteer and observer.

If animals or children are observed, devices can be synced in real time where one device mimics the actions done on another device. Button presses can be indicated by highlighting the buttons and playing a short sound clip

which allows inexperienced observers to view the coding of behavior in an actual setting, similar to the "spectator mode" in computer games [37].

While the analysis of the observation in terms of supporting or rejecting hypotheses should probably be avoided to avoid observer bias, "on-the-fly" statistical reliability analyses can be useful. If two observers shadow the same subject, their observations can be compared and deviations can be discussed in the situation itself. Taking different perspectives into account, this might at least give some idea of the reliability of the observations and still be able to improve them, esp. if no recoding based on video data can be done.

If it is difficult to identify behaviors during the observation, quickly available images, videos, or audio files within the app of different instances of the same behavior might refresh memory and reduce ambiguity. Examples should also include similar but different behaviors with explanations of the differences, e.g., to provide some context for what is regarded as a "stop" in a museum (see discussion in [2]). Videos can be created by recording a volunteer/animal first. A button can provide quick access to the examples by changing the function of all other buttons to show an example screen instead of logging the behavior. These examples can even make it possible for laypeople to code the observed behaviors correctly (see "Subject"). For new behavior, connectivity (e.g., via messages) can be used to let different observers exchange their notes on coding ambiguous behavior. Hints about the observation can also be shared while the observation is ongoing.

To support objectivity, CADO apps can provide checklists for setting up the observation, rules for the observation, help files, tutorials for the app or the specific observation, and much more. Checklists with completion checks that reset after each observation period can be used. This would require observers to check and confirm that they have completed the necessary steps and thus avoid variation in the observation. Likewise information about ethical issues and best practices can be made available.

3) Communication among observers for coordination

If more than one observer is used, connectivity can be used to sync the start times of the observations. Connectivity in this case is achieved either by establishing a connection between the devices directly or by connecting the devices to a time server (Internet Time Service).

Connectivity, esp. in combination with location sensitivity, can be used to guide multiple observers to good positions, to coordinate time and location sampling [4], or to let observers organize themselves via chat or voice messages. Chat and voice messages among the observers should be recorded as well to prevent misuse. Connectivity can also be used to randomly assign observers to sampling units (see [4]). If quotas are used, connectivity can aid in adhering to these quotas. The demographics of all finished observations can be matched against the quota and the desired targets would be presented to the observers. Coordinating multiple observers to shadow subjects more effectively can also be useful. Instead of one observer following a subject the whole time, the subject could be passed to the next observer. A short shared description of the subject would clarify the target if photos cannot be used. In the same way, multiple observers can also code

the complex behavior of one person by dividing up the behaviors to code, e.g., one focusing on talk, the other on movement, the third on gaze. Existing data on behavior frequency, e.g., based on video coding, can be used to divide the behaviors in a way that allow for direct observation coding. The virtual interface then displays the respective buttons for each observer.

One step further, gaze sensors of smartphones can be used. Some smartphones are able to determine when the user looks at the screen to automatically scroll text or to stop videos. However, gaze sensors can be used to coordinate observers. As no behavior can be visually observed when an observer looks at the screen, a look at the screen can automatically signal another observer to take over the observation to ensure a more complete coverage of observed behaviors.

4) *Communication with the lead researcher*

If the application is able to import/export the observation interface and other information, then it becomes possible to easily update the observation schemes by sending the data via eMail, for example, or exchanging it between the devices directly (e.g., via WiFi or Bluetooth). Incremental updates (delta updates) keep the bandwidth to a minimum, which is relevant in field conditions.

Connectivity allows sending the observation data to backup servers immediately when a connection is available. With an application that aggregates the data in a dashboard (similar to content management systems), the supervisor can examine the data as it is collected.

While eMails can be used to exchange data (see [3]), questions about ambiguous behavior can also be asked. For example, if simple human behavior is coded, a volunteer/second observer can record the observer repeating the behavior. The supervisor can then decide how to code this behavior and distribute that information to all observers.

IV. DISCUSSION

The six different but interconnected factors of the activity theory for mobile learning — tool, subject, objective, control, context, and communication — point to ways to improve observation with digital devices. However, it is likely that not all improvements are equally beneficial, nor beneficial for all observations.

This said, probably the greatest possible potential is to make use of the ubiquity of mobile devices and get laypeople to participate in scientific observations. Smartphones and touchscreen-based pocket computers like the iPod touch are widely available. There are already Apps used privately for life-logging (e.g., [38]) or self-observation (e.g., Ethnography Apps like "MyServiceFellow" and "Over The Shoulder", see [39]). Laypeople are already continually observing their environment or themselves, often either informing scientists about their discoveries (e.g., astronomers), working in close cooperation with ethnographers, or using the data themselves (e.g., the quantified self movement). If the App is less a specialist piece of software available only to a few, but more user-friendly and self-explanatory, more quantitative observations can be done by laypeople as well. Despite usually being a complex one-shot task, a well-designed CADO app can do for observations what the Internet and Mechanical Turk has done for surveys — with vast benefits for science.

However, there are potential downsides to using commercial, privately owned devices. Observers with different devices might increase error variance, although minimum requirements in display size and processing speed should keep them in check, and it is coding behavior in the same way that counts, not having the same devices. Most smartphones override the current app when a call comes in, but this can be prevented by using the airplane mode (if communication is not required) or by using a "do not disturb" mode to block calls (if the observer agrees). Smartphones are usually more common with younger people of higher socio-economic status. However, these people might also be the ones who are more likely to work as observers. After all, if recruited via Mechanical Turk, laypeople as observers require a certain amount of technical equipment to be on Mechanical Turk in the first place. Network coverage can be difficult during observations in some areas (e.g., parts of the developing world), so data should be stored locally on the device first. The device with its personal information and apps might also be a source of distraction. However, this can be assessed if the app logs when it is closed or moves to the background.

Given that the collected data is about other people, there are ethical and legal implications. However, some surveys also collect data about other people (e.g., surveys about domestic violence) or the social environment (e.g., surveys about the work climate), but no video data of other people is produced. Possible observers have to be briefed and recruited carefully. The potential to use the devices directly for training, esp. in handling problematic situations, is useful here.

The descriptions and the example focused on smartphone sized devices. However, almost all functions can be carried out as well on tablet sized devices. For iOS, projects can be easily converted or developed for iPhone/iPod touch and iPad at the same time. Whereas tablets offer more screen real estate, they are more obtrusive in many settings. They are larger, easier to see and more noticeable because few people carry tablets with them. Thus, there are good reasons to use smaller handhelds. However, this depends on the needs of the specific observation.

While the possible features might sound like pipe dreams or science fiction, they are feasible from a technical perspective. The iOS and its Software Development Kit allows for the implementation of these functions. Ref. [40] discusses different ways of implementing applications for research purposes. He looks at general-purpose programs and specific-experiment programs, and argues for a hybrid approach, using object-oriented programming and XML objects to select and arrange those objects. A similar approach can be used for such an observation app: Using objects (incl. variables and functions) and an interface to create and arrange these objects, depending on the needs of the observation.

Also, the code can be adapted to specific needs by adding new objects, adding variables or functions to existing objects, or by subclassing them. The specific observation setup can be exported/imported as observational modules, which include instructions, interface, training files, images, videos, etc.

However, some of these functions require a large programming effort. While many features are waiting to be used by commercial providers of mobile observation

software, there is also huge potential for non-commercially developed apps. There is a long tradition of scientific participation, and there are good reasons for scientists to develop new tools, or to be closely involved in their development (see, e.g., [41]; [42]). For example, some research needs are too specific for commercial software [41]. Science grows and improves with new tools [41], it is less constrained by other people's conceptions and allows for more creative solutions ([41]; [42]). The scientific imagination also encourages innovation [43] (also one reason for the present paper), and it allows for better cross-disciplinary communication, even if a professional programmer or company does the actual development and programming [42]. If enough scientist-programmers develop their observation tools this might stimulate the evolution of apps. An Open Source App as basis with modules to augment it would be a good first step. While such a project can become unwieldy quickly — the technology is advanced and stable enough to work with.

Technology that is usable for observations will develop further, and new, interesting developments are already on the horizon. This might lead to reluctance to develop apps for current, widely distributed technology. After all, glasses with integrated optical head-mounted display are close to coming on the market or available already (e.g., Google Glass). These glasses allow observers to keep their eyes on the subject while coding the behavior. However, it essentially only changes the interface, not the underlying functions. In fact, most of the potential discussed can be used in this case as well. Mobile media is not constrained only to smartphones, but can be applied to other devices as well.

Thus it makes sense to start realizing the underused potential of mobile media for scientific observations. Science improves with better tools [41] and with ingenuity that “leads to new means of making more valid comparisons” [4]. Whether they will actually fulfill this potential is beyond the scope of this article. However, to find out and possibly improve observation studies, it is high time to implement these tools.

ACKNOWLEDGMENT

I thank Lars Kobbe for being the first in the department to program a simple observation tool for a Fujitsu-Siemens Pocket Loox using HTML and JavaScript. It started my imagining what is possible with these devices. I also thank Daniela Bauer for trying out my prototype in a small exhibition in Norway, and Jürgen Buder for his valuable feedback on the first version of this article. The author declares no conflict of interest.

REFERENCES

- [1] R. D. Ray, J. M. Ray, D. A. Eckerman, L. M. Milkosky, and L. J. Gillins, “Operations Analysis of Behavioral Observation Procedures: A Taxonomy for Modeling in an Expert Training System,” *Behavior Research Methods*, vol. 43, no. 3, pp. 616–634, Jul. 2011. <http://dx.doi.org/10.3758/s13428-011-0140-6>
- [2] S. S. Yalowitz and K. Bronnenkant, “Timing and Tracking: Unlocking Visitor Behavior,” *Visitor Stud.*, vol. 12, no. 1, pp. 47–64, Apr. 2009. <http://dx.doi.org/10.1080/10645570902769134>
- [3] J. M. Vivoda and D. W. Eby, “Using Personal Digital Assistants (PDAs) for the Collection of Safety Belt Use Data in the Field,” *Behavior Research Methods*, vol. 38, no. 1, pp. 158–164, 2006. <http://dx.doi.org/10.3758/BRM.38.1.158>
- [4] E. J. Webb, D. T. Campbell, R. D. Schwartz, and L. Sechrest, *Unobtrusive Measures*, 1966 ed. Thousand Oaks, California: Sage Publications, Inc., 1966.
- [5] K. Ellis-Davies, E. Sakkalou, N. C. Fowler, E. E. Hilbrink, and M. Gattis, “CUE: The Continuous Unified Electronic Diary Method,” *Behavior Research Methods*, vol. 44, no. 4, pp. 1063–1078, May 2012. <http://dx.doi.org/10.3758/s13428-012-0205-1>
- [6] A. L. Sillars, “Behavioral Observation,” in *Studying Interpersonal Interaction*, no. 11, B. M. Montgomery and S. Duck, Eds. New York: The Guilford Press, 1991, pp. 197–218.
- [7] M. S. Goodwin, W. F. Velicer, and S. S. Intille, “Telemetric Monitoring in the Behavior Sciences,” *Behavior Research Methods*, vol. 40, no. 1, pp. 328–341, 2008. <http://dx.doi.org/10.3758/BRM.40.1.328>
- [8] A. Agliati, A. Vescovo, and L. Anolli, “A New Methodological Approach to Nonverbal Behavior Analysis in Cultural Perspective,” *Behavior Research Methods*, vol. 38, no. 3, pp. 364–371, 2006. <http://dx.doi.org/10.3758/BF03192789>
- [9] O. H. Maclin and M. K. Maclin, “Coding Observational Data: A Software Solution,” *Behavior Research Methods*, vol. 37, no. 2, pp. 224–231, 2005. <http://dx.doi.org/10.3758/BF03192690>
- [10] R. Bakeman, V. Quera, and A. Gnisci, “Observer Agreement for Timed-Event Sequential Data: A Comparison of Time-Based and Event-Based Algorithms,” *Behavior Research Methods*, vol. 41, no. 1, pp. 137–147, Feb. 2009. <http://dx.doi.org/10.3758/BRM.41.1.137>
- [11] D. vom Lehn, C. Heath, and J. Hindmarsh, “Video Based Field Studies in Museums and Galleries,” *Visitor Stud. Today*, vol. 5, no. 3, pp. 15–23, 2002.
- [12] J. Held and T. Manser, “A PDA-Based System for Online Recording and Analysis of Concurrent Events in Complex Behavioral Processes,” *Behavior Research Methods*, vol. 37, no. 1, pp. 155–164, 2005. <http://dx.doi.org/10.3758/BF03206410>
- [13] H. Cartier-Bresson, *The Decisive Moment*. New York: Simon and Schuster, 1952.
- [14] S. Kahng and B. A. Iwata, “Computerized Systems for Collecting Real-Time Observational Data,” *Behavior Research Methods, Instruments, & Computers*, vol. 31, no. 2, pp. 253–261, 1998.
- [15] J. Tapp, R. Ticha, E. Kryzer, M. Gustafson, M. R. Gunnar, and F. J. Symons, “Comparing Observational Software with Paper and Pencil for Time-Sampled Data: A field test of Interval Manager (INTMAN),” *Behavior Research Methods*, vol. 38, no. 1, pp. 165–169, 2006. <http://dx.doi.org/10.3758/BF03192763>
- [16] A. Sarkar, A. Dutta, U. Dhingra, P. Dhingra, P. Verma, R. Juyal, R. E. Black, V. P. Menon, J. Kumar, and S. Sazawal, “Development and Use of Behavior and Social Interaction Software Installed on Palm Handheld for Observation of a Child’s Social Interactions with the Environment,” *Behavior Research Methods*, vol. 38, no. 3, pp. 407–415, 2006. <http://dx.doi.org/10.3758/BF03192794>
- [17] D. J. White, A. P. King, and S. D. Duncan, “Voice Recognition Technology as a Tool for Behavioral Research,” *Behavior Research Methods, Instruments, & Comput.*, vol. 34, no. 1, pp. 1–5, 2002. <http://dx.doi.org/10.3758/BF03195418>
- [18] “iOS Human Interface Guidelines,” *apple.com*, 2014. [Online]. Available: <https://developer.apple.com/library/iOS/documentation/userexperience/conceptual/mobilehig/MobileHIG.pdf>. [Accessed: 01-Jul-2014].
- [19] B. Millet, S. Asfour, and J. R. Lewis, “Selection-Based Virtual Keyboard Prototypes and Data Collection Application,” *Behavior Research Methods*, vol. 41, no. 3, pp. 951–956, Aug. 2009. <http://dx.doi.org/10.3758/BRM.41.3.951>
- [20] J. Taylor, M. Sharples, C. O’Malley, G. N. Vavoula, and J. Waycott, “Towards a Task Model for Mobile Learning: a Dialectical Approach,” *Int. J. of Learning Technology*, vol. 2, no. 2, pp. 138–158, 2006. <http://dx.doi.org/10.1504/IJLT.2006.010616>
- [21] M. Sharples, J. Taylor, and G. Vavoula, “A Theory of Learning for the Mobile Age,” in *The Sage Handbook of Elearning Research*, R. Andrews and C. Haythornthwaite, Eds. London: Sage, 2007, pp. 221–247.
- [22] D. Froberg, C. Göth, and G. Schwabe, “Mobile Learning Projects – A Critical Analysis of the State of the Art,” *J. of Computer As-*

PAPER
THE POTENTIAL OF COMPUTER-ASSISTED DIRECT OBSERVATION APPS

- sisted Learning*, vol. 25, no. 4, pp. 307–331, Jul. 2009. <http://dx.doi.org/10.1111/j.1365-2729.2009.00315.x>
- [23] B. Fang, “From Distraction to Engagement: Wireless Devices in the Classroom,” *EDUCAUSE Quarterly*, vol. 32, no. 4, pp. 4–9, 2009.
- [24] E. Klopfer, K. Squire, and H. Jenkins, “Environmental Detectives: PDAs as a Window into a Virtual Simulated World,” presented at the IEEE International Workshop on Wireless and Mobile Technologies in Education, 2002 (WMTE'02), 2002, pp. 95–98.
- [25] “PASPORT Sensors and Probes,” *pasco.com*. [Online]. Available: <http://www.pasco.com/sensors/>. [Accessed: 01-Jul-2014].
- [26] S. Reimers and N. Stewart, “Using Adobe Flash Lite on Mobile Phones for Psychological Research: Reaction Time Measurement Reliability and Interdevice Variability,” *Behavior Research Methods*, vol. 40, no. 4, pp. 1170–1176, 2008. <http://dx.doi.org/10.3758/BRM.40.4.1170>
- [27] D. R. Thorne, D. E. Johnson, D. P. Redmond, H. C. Sing, G. Belenky, and J. M. Shapiro, “The Walter Reed Palm-Held Psychomotor Vigilance Test,” *Behavior Research Methods*, vol. 37, no. 1, pp. 111–118, 2005. <http://dx.doi.org/10.3758/BF03206404>
- [28] E. Hand, “People Power,” *Nature*, vol. 466, Aug. 2010. <http://dx.doi.org/10.1038/466685a>
- [29] K. Werbach and D. Hunter, *For the Win*. Philadelphia: Wharton Digital Press, 2012.
- [30] S. Yalowitz and A. Ferguson, “Sharks: Myth and Mystery,” Monterey Bay Aquarium, 2006.
- [31] J. Gerich, “Visual Analogue Scales for Mode-Independent Measurement in Self-Administered Questionnaires,” *Behavior Research Methods*, vol. 39, no. 4, pp. 985–992, 2007. <http://dx.doi.org/10.3758/BF03192994>
- [32] D. Wessel and E. Mayr, “Potentials & Challenges of Mobile Media in Museums,” *Int. J. of Interactive Mobile Technologies*, vol. 1, no. 1, pp. 32–39, 2007.
- [33] “Digital Pen,” *wikipedia.org*. [Online]. Available: http://en.wikipedia.org/wiki/Digital_pen. [Accessed: 01-Jul-2014].
- [34] “Star Walk,” *vitotechnology.com*. [Online]. Available: <http://vitotechnology.com/star-walk.html>. [Accessed: 01-Jul-2014].
- [35] K. L. Harris, “Content Analysis in Negotiation Research: A Review and Guide,” *Behavior Research Methods, Instruments, & Comput.*, vol. 28, no. 3, pp. 458–467, 1996. <http://dx.doi.org/10.3758/BF03200525>
- [36] “The MP3 Experiments,” *improveeverywhere.com*. [Online]. Available: <http://improveeverywhere.com/missions/the-mp3-experiments/>. [Accessed: 07-Jan-2014].
- [37] “Electronic Sports,” *wikipedia.org*. [Online]. Available: http://en.wikipedia.org/wiki/Electronic_sports#Spectator_mode. [Accessed: 01-Jul-2014].
- [38] “Momento,” *momentoapp.com*. [Online]. Available: <http://www.momentoapp.com>. [Accessed: 01-Jul-2014].
- [39] M. Gotta, “Your Sensored Life: An Expanded View of Quantified Self,” *gartner.com*, 2014. [Online]. Available: <http://blogs.gartner.com/mike-gotta/your-sensored-life-an-expanded-view-of-quantified-self/>. [Accessed: 01-Jul-2014].
- [40] P. Dixon, “A Hybrid Approach to Experimental Control,” *Behavior Research Methods*, vol. 41, no. 3, pp. 615–622, Aug. 2009. <http://dx.doi.org/10.3758/BRM.41.3.615>
- [41] W. K. Beagley, “Why We Need More Psychology Programmers/EL Knife, a Data Utility for Transforming Spreadsheets,” *Behavior Research Methods, Instruments, & Comput.*, vol. 33, no. 2, pp. 97–101, 2001. <http://dx.doi.org/10.3758/BF03195353>
- [42] P. Guo, “Helping Scientists, Engineers to Work up to 100 Times Faster,” *Commun. ACM*, vol. 56, no. 10, p. 12, Oct. 2013. <http://dx.doi.org/10.1145/2507771.2507775>
- [43] O. Pineño, “ArduiPod Box: A Low-Cost and Open-Source Skinner Box Using an iPod Touch and an Arduino Microcontroller,” *Behavior Research Methods*, Jun. 2013.

AUTHORS

D. Wessel is with the Leibniz-Institut für Wissensmedien, Schleichstrasse 6, 72076 Tuebingen, Germany (e-mail: d.wessel@iwm-kmrc.de).

Submitted 10 October 2014. Published as resubmitted by the author 25 January 2015.