OPEN ACCESS

University of Dundee

University of Dundee

End-user development in social psychology research

Rough, Daniel; Quigley, Aaron

Link to publication in Discovery Research Portal

Download date: 07. Jul. 2020

End-User Development in Social Psychology Research: Factors for Adoption

Daniel Rough
SACHI Research Group
University of St Andrews
Email: djr53@st-andrews.ac.uk

Aaron Quigley
SACHI Research Group
University of St Andrews
Email: aquigley@st-andrews.ac.uk

Abstract—Psychology researchers employ the Experience Sampling Method (ESM) to capture thoughts and behaviours of participants within their everyday lives. Smartphone-based ESM apps are increasingly used in such research. However, the diversity of researchers' app requirements, coupled with cost and complexity of their implementation, has prompted end-user development (EUD) approaches. In addition, limited evaluation of such environments beyond lab-based usability studies precludes discovery of factors pertaining to real-world EUD adoption. We first describe the extension of Jeeves, our visual programming environment for ESM app creation, in which we implemented additional functional requirements, derived from a survey and analysis of previous work. We further describe interviews with psychology researchers to understand their practical considerations for employing this extended environment in their work practices. Results of our analysis are presented as factors pertaining to the adoption of EUD activities within and between communities of practice.

I. INTRODUCTION

The Experience Sampling Method (ESM) is a methodology employed primarily in psychology and clinical research to survey participants in-the-moment, in their natural contexts [1], [2]. ESM has a number of advantages over retrospective surveys; primarily, it has high ecological validity as participants are assessed in their natural environments, as opposed to a lab or clinic. Further, it minimises recall bias, as participants' experiences are captured in situ. As a recent example of such advantages, Lenaert et al. demonstrate how ESM can be used to identify fluctuating emotions in patients with Acquired Brain Injury, to improve researchers' understanding and consequent treatment [3].

Although smartphones are an ideal medium for delivery of ESM surveys, previous work has recognised that researchers face significant programming barriers to adoption of smartphone-based ESM, given that the necessary software development skills are often outside their area of expertise [4], [5]. Short of defaulting to paper-based methods, researchers must rely on professional support to develop custom apps for their studies. As well as the expense of this approach, bespoke apps are inflexible to diverse, dynamic requirements of researchers and their participants. With respect to these issues, end-user development (EUD) presents a possible solution [6].

EUD tools for ESM apps have previously been developed in response to this need, which we survey in Section II. However, despite calls to investigate the socio-technical aspects of EUD introduction, there is little research on the impact of such technology in working practices. Thus we ask the following research question: *What factors influence the adoption of technology for psychology researchers to develop experience sampling smartphone apps?*

In this paper we make two research contributions, centred on the design of an extensible EUD tool for ESM app creation (hereby referred to as an EUD-ESM tool). We first survey the potential utility of ESM apps and derive requirements towards facilitating social psychology research, and we describe the supporting design decisions implemented in the novel extension of *Jeeves*, our existing EUD-ESM tool [7].

As a second contribution, we address the paucity of research into adoption requirements for EUD in working practices, through qualitative analysis of psychology researchers' perceptions of *Jeeves*, and EUD practice in general, obtained through semi-structured interviews. Analysing our results with respect to current models of technology acceptance, we discuss the necessary factors for adoption of an EUD-ESM tool in social psychology, and how these relate to general "public-outward EUD", where one community of end-users develops for another [8], thus broadening the implications of this study.

II. RELATED WORK

Research relevant to the goals of our work covers three areas - potential utility of ESM, existing EUD-ESM tools, and adoption of EUD in professional practice. In a traditional model of ESM development, researchers express their requirements to a programmer, who creates a bespoke app for that particular study. Participants install the app and return to researchers after the study period. In our proposed update to this model (Figure 1), these stakeholders have separate development roles, represented by the three levels of tailoring defined by Mørch [9]. The programmer's role is that of a *meta-designer*, who creates the components necessary for the researcher through **extension**. Using these components, the researcher can **create** or **modify** apps, with behaviour that can be **customised** by study participants. Beyond development paradigms that are usable by non-programmers, EUD encompasses socio-technical aspects of software. Fischer's
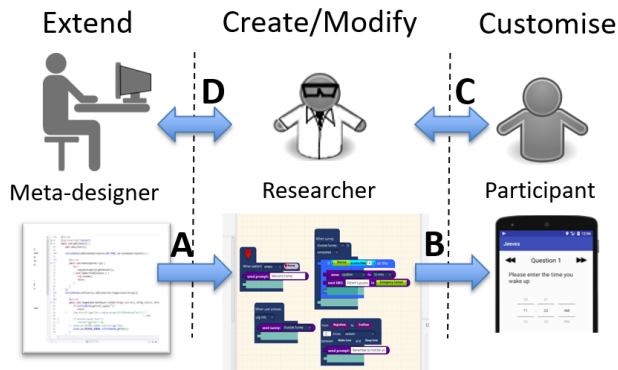
Fig. 1. Our meta-design model of ESM app development, where stakeholders collaborate and perform EUD activities "during use"

meta-design framework acknowledges that development is a continuous effort that involves ongoing collaboration [10].

A. Potential Utility of ESM

A review of recent literature highlights two application areas of repeated, longitudinal assessment inherent in ESM:

1) In **research**, to investigate how participant variables fluctuate over time and in different contexts
2) In **practice**, to allow participants to independently monitor aspects of their mental health

Previous work has surveyed the benefits of ESM to behavioural research [4], [5], and the process of in-situ, longitudinal assessment and intervention has been thoroughly discussed as an asset to participants' health in practical applications [11], [12]. Additionally, general mobile health (mHealth) interventions have been surveyed [13], providing evidence-based recommendations for features relevant to ESM, such as self-monitoring. With these applications in mind, our related work summarises potential beneficial features of smartphone ESM to both researchers, and participants themselves.

Objective Context Triggering: Automated capture and inference of objective sensor data can support the delivery of assessments, interventions, or reminders at ideal times, introducing new possibilities for psychology research. For example, contextual sensing can be used to predict the "interruptibility" of participants, to ensure surveys are sent at minimally intrusive moments [14]. In a medical context, participants' mood was inferred from sensor values to deliver tailored feedback in an intervention for depression [15]. Location-based reminders have been endorsed as a useful feature in behavioural interventions [16], and feedback on smartphone-sensed activity can also promote self-awareness [17].

Subjective Context Triggering: A deliberate distinction has been made between *objective* and *subjective* context triggering, whereby the latter refers specifically to an app performing actions based on participants' subjective, self-reported information. For example, "Ecological Momentary Interventions" (EMIs) can support positive behaviour change through in-the-moment, tailored delivery of prompts or coping strategies to participants, without need for direct researcher involvement [18]. Further, medication reminders based on participants' self-reported administration have been perceived as useful in user-centred design studies [11], [19].

Two-Way Feedback: In practical applications of ESM, automated feedback is not a substitute for human contact. Allowing researchers to directly prompt lapsing participants, and participants to report issues to researchers, has been identified as useful for studies in individuals with mental illness, for example [20]. Participatory sensing literature suggests that feedback from researchers could act as a non-monetary incentive mechanism, motivating participants as active contributors to a study [21]. Indeed, employing participatory design as part of an ESM app could enable researchers to immediately address study design issues through direct participant feedback [22].

Preference Tailoring: Just as meta-designers cannot predict the requirements of researchers, researchers cannot entirely predict participants' needs and characteristics. Participants with hectic schedules support manually tailoring survey prompt times [16], while prompts that are delivered excessively or at inconvenient times are likely to frustrate and result in non-compliance [23]. In an empirical study testing this hypothesis, allowing participants to personalise their sampling times significantly increased their responsiveness to surveys [24]. Other research has also proposed the use of personalised survey schedules to increase compliance [25].

B. Existing EUD-ESM Tools

Responding to the programming barrier faced by psychology researchers, a number of tools exist, both as research projects and proprietary systems, to facilitate ESM development. Table I provides a summary of recent and prominent efforts. We searched the ACM, IEEE and Scopus digital libraries using the terms 'experience sampling', 'ecological momentary', 'end-user development', 'end-user programming', and 'smartphone' to derive tools in research. Further, given the prevalence of such tools in the commercial domain, a standard Google search was used with the search terms listed above, in an attempt to find proprietary examples. Finally, Conner's resource on ESM creation tools was used to identify further efforts [26]. The table lists these tools in relation to the four potential ESM features previously identified, which are explored in the following section.

1) Objective Context: Tools that enable specification of sampling schedules based on objective context present limited funtionality. While the proprietary platforms *LifeData*, *MovisensXS* and *EthicaData* enable the GPS tagging of self-reports, none enable the researcher to trigger based on this location, or indeed other sensors. *AWARE* and *Ohmage* are open-source software platforms that could be programmed to enable sensor triggering, and *EthicaData* provides an API through which developers can create and link their own trigger functionality, but none include this functionality in their available state. Of the platforms that do, *Sensus* is the most diverse in its triggering capabilities, supporting external devices as well as on-device sensors. *PartS* additionally provides a visual interface for specifying objective context triggers.

Platform Name	Objective Trigger	Subjective Trigger	Preference Tailoring	Two-Way Feedback
SurveySignal [28]	✗	✗	✗	✗
LifeData [29]	✗	✗	✗	✗
MovisensXS [30]	✗	✗	✗	✗
PsychLog [31]	✗	✗	✗	✗
EthicaData [32]	○	○	○	✗
PACO [33]	✓	○	○	○
AWARE [34]	○	○	○	○
Ohmage [35]	○	○	○	○
Sensus [36]	✓	○	○	○
PartS [21]	✓	✗	✗	✓

2) Subjective Context: None of the tools surveyed allow for subjective context triggering, such as performing actions that are contingent on participants' responses to surveys. Such functionality would be necessary in order to provide tailored intervention feedback to participants. However, self-report data in all reviewed tools cannot be interpreted by the apps themselves; any intervention would need to be initiated by the researcher manually after reviewing participants' data.

3) Preference Tailoring: Functionality for tailoring to the characteristics of individual participants has also not been implemented in the tools of Table I. With *ESP*, one of the first examples of electronic ESM, participants used palmtop computers that were manually programmed to account for their waking and sleeping times [27]. This is burdensome for researchers, and inflexible to changes in participants' schedules. *Ohmage* and *PartS* allow participants to set their own reminders but the researcher has no ability to add other customisations.

4) Two-way Feedback: Communication functionality is understandably limited in existing tools. Given the requirements of anonymity and consistency inherent in ESM research studies, direct researcher-participant communication has potential ethical implications. Moreover, biased communication delivered to some participants, but not others, could bring the validity of collected data into question. Some tools enable study information messages to be sent to all participants, but with no way for these participants to provide feedback, or to have individual interactions. *PartS*, as a participatory sensing platform, is an exception to this, supporting two-way communication as a motivational incentive for participants.

C. Adoption of EUD in Practice

The prevalence of lab-based usability studies in the evaluation of EUD tools is contrasted by the lack of research into their real-world utility [37], a disparity recognised within HCI as a whole [38]. EUD evaluations are largely focused on the development paradigm and how users' mental models of programming tasks affect the usability of particular paradigms. However, the development of *useful* EUD environments requires knowledge of who potential end-users are, their goals
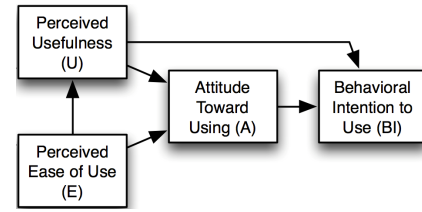

Fig. 2. The Technology Acceptance Model, adapted from [41]

and motivations, and how such environments could fit with current working practices.

As an example of this, recent work by Namoun et al. discusses design implications for mobile EUD from results of surveys and focus groups [39]. The qualitative methods employed, and subsequent data analysis, provide detailed design implications for informing future work. Tetteroo et al. investigate socio-technical factors of introducing EUD in a clinical setting [40], providing deployment guidelines that transcend factors of usability in this domain.

In work outside of EUD, models have been derived that predict the adoption success of general technology, including the Technology Acceptance Model (TAM) [41], illustrated in Figure 2, and the Unified Theory of Acceptance and Use of Technology (UTAUT) [42]. Core factors of both these models are that users' intention to adopt technology is influenced by both its usability, and usefulness. Although these models are applicable to a variety of software, EUD is unusual, in that it involves end-users in a non-traditional role as developers. A question of this research is thus whether these factors are sufficient to capture acceptance of EUD tools in practice.

III. TOOL EXTENSION

For our required evaluations, we updated our existing EUD-ESM tool, *Jeeves*, which employs a visual programming paradigm, specifically a blocks-based approach similar to that of the App Inventor [43] environment. An evaluation of this paradigm showed that non-programmers did not significantly differ in task time or error rate from programmers, and that both perceived *Jeeves* as usable across many dimensions [7]. With an aim to develop *Jeeves* into a tool that supports meta-design, we derived and implemented features to extend and enhance its current functionality, rethinking the desktop interface and smartphone app as "software shaping workshops" as proposed in previous meta-design literature [10].

At the researcher's level, our extended version of *Jeeves* acts as a *system workshop* where researchers create and modify apps to suit their research question or participant group. The app itself acts as an *application workshop* whereby participants can potentially customise their experience to fit their everyday lives. The modular implementation of *Jeeves*, where functionality is composed of different block types, supports simple extension by meta-designers to cope with the changing requirements of researchers. This section discusses our implemented extensions and features, which we performed as part of Figure 1A.
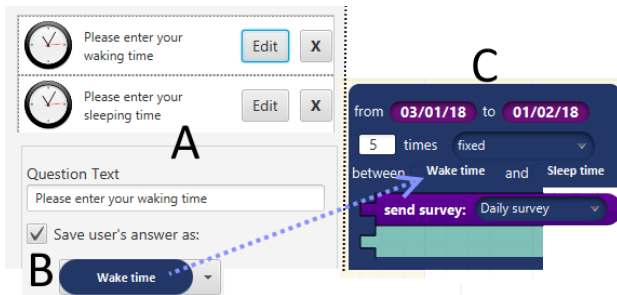
Fig. 3. App behaviour can be customised by storing answers to survey questions in attribute blocks

Fig. 4. Objective context triggers can be set up on semantic locations, and combined with other participant states

A. User Data Pane

A new pane was implemented to provide a visual interface to real-time, incoming participant data, and to allow feedback to be sent and received (Figure 1C). This "User Data Pane" consists of a simple GUI where the number of surveys completed and missed by each participant can be viewed. Additionally, **two-way feedback** is afforded through a messaging widget that displays messages sent to and from individual participants via their app.

B. User Attributes

User Attributes were realised as a series of blocks that represent state characteristics of an individual participant, analogous to program instance variables. *Jeeves* contains a new pane for attribute creation, and by using these attributes in a specification, researchers can implement participant-specific app behaviour, as part of their EUD activities (Figure 1B), enabling **subjective context triggering** to be incorporated.

Apps must incorporate participant preferences. For example, participants could select individual waking and sleeping times, or specific locations to receive prompts. In our extension, attribute values are set by survey responses, affording participants **preference tailoring** functionality by answering assigned questions. An example of the process is demonstrated in Figure 3. In this example, the survey is designed to ask participants what their waking and sleeping times are (A). Their answer to the "waking time" question is saved into the "Wake time" attribute (B), which is then used to tailor the schedule of a time-based trigger (C), along with a corresponding "Sleep time" attribute. Participants can then customise a variety of attributes. For example, a survey question can prompt the participant for a GPS location, to customise a geofencing trigger, an example of which is shown in Figure 4.

C. Context-Sensitive Triggering

The previous version of *Jeeves* allows creation of triggering functionality based on a change in location or acceleration. To allow for richer context-sensitive sampling, our updated *JeevesAndroid* smartphone app employs Google's Activity Recognition API, allowing a variety of activities to be used as triggers. Further, the Google Places API allows geofences to be set at *semantic* locations specified by a participant, such
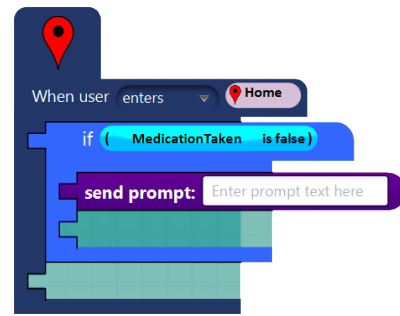
as their home or workplace. An example of this **objective context triggering** is shown in Figure 4, where a participant's specified "Home" location, and their current medication state, are combined to trigger a medication reminder prompt.

IV. FACTORS FOR ADOPTION

Following extensions in *Jeeves* with features perceived to be particularly beneficial to ESM studies, we sought to investigate whether these features would be influential in its adoption by social psychology researchers, and further, whether adoption would be contingent on other factors previously unconsidered. Although we separately evaluated the usability of these features, the "perceived ease-of-use" construct of the TAM in Figure 2 is not the focus of this work, as previously discussed. Instead, "perceived usefulness" was evaluated through qualitative research with potential end-users.

A. Interviews

Semi-structured interviews were conducted with five social psychology researchers at our university, recruited through personal email requests. Their research areas and relevant participant cohorts were sufficiently diverse to obtain a range of considerations for adoption of *Jeeves*. Interviews took place at the researchers' location of choice, lasted approximately 45 minutes, and were organised around three questions:

- **Current practices**: What are researchers' existing practices in the study of participants, their benefits and drawbacks?
- **Technology use**: What are researchers' perceptions and experiences of technology in their current practices, and what motivates them to use such technology?
- **Initial impressions**: After being shown an example specification created with *Jeeves*, can researchers envision further applications?

Thematic coding linked researchers' feedback to factors in the aforementioned existing models of technology acceptance [42]. The relevant factors are *perceived usefulness* (which is further divided into *perceived potential*, *initial requirements* and *participant requirements*) and, exclusive to the UTAUT, *facilitating conditions*.

B. Perceived Potential

The perceived potential benefits of *Jeeves* were elicited from researchers following demonstration of the tool's capabilities. These were primarily related to their difficulty in conducting ecologically valid research, described as follows, and summarised in Table II. These potential benefits are specific to ESM apps, but broad motivations of **saved time**, **functional quality** and **participant quality** are applicable to all public-outward EUD.

Jeeves could enable compliance monitoring

Compliance remains a major issue for studies where participants are required to maintain active participation outside a lab environment, and researchers explained that significant time was wasted through participant dropout. P4 explained how tracking the number of completed surveys, as well as the time taken to complete these surveys, could be used to motivate participants with additional financial compensation:

"if you could have a mechanism I guess of...you know recording if you complete all these bits we'll put you in a prize draw...and you can see if they've done the whole survey in 10 seconds"

The ability to manually prompt participants to complete surveys was also valued by P3, who explained his use of *Qualtrics* software in order to avoid sending unnecessary compliance emails:

"[Qualtrics] keeps track of who's not responded yet so you can send up a follow-up email to only those who've not responded...allows you to interact with your participant pool"

Jeeves could eliminate recall inaccuracy

Alleviating recall bias and ecological validity issues caused by lab-based data collection was the most significant perceived benefit. Researchers described how a longitudinal experience sampling study would help to alleviate the memory biases that occur due to the time lapse between an event and its reporting, and collect data of a quality previously unattainable.

"The closer you can get to the actual event, and treating each event as a unit...we get a bit closer to the raw experience itself in some way." (P3)

Jeeves could enable capture of contextual information

Context-contingent assessments were perceived as highly desirable in practical applications of *Jeeves*. P5 explained a study she hoped to conduct investigating contextual influences on participants' mindfulness, such that location-based triggers would allow accurate, quality information to be collected:

"That's something we want to develop. 'Where are you? Are you in your bedroom, meditating? Are you outside?' That could be extremely useful"

In situ, repeated self-reports were acknowledged as potentially disruptive. However, researchers suggested that contextual triggers could minimise the number of unnecessary interruptions. P4 described a further potential application of the location trigger functionality to minimise interruptibility:

"If you had the location, you could, as soon as they left...at that point it's appropriate right now to ask them

TABLE II
PERCEIVED POTENTIAL OF JEEVES, FROM INTERVIEW FEEDBACK

Specific Benefit	Broad Motivation	Required Feature(s)
Researchers can prompt compliance	Saved Time	Two-Way Feedback
Researchers acquire in-the-moment data	Functional Quality	
Researchers reduce recall inaccuracy	Functional Quality	Objective Triggering Subjective Triggering
Time saved through remote research	Saved Time	
Participants engage in self-monitoring	Participant Quality	Preference Tailoring
Intervention for sensitive participants	Participant Quality	Subjective Triggering Two-way Feedback

what happened...it's going to be fresh in their mind without interrupting their experience, that'd be great"

Jeeves could save time over traditional data collection

The cross-sectional nature of most research was apparent, and researchers were explicit about the disadvantages of this approach. P4's research of experiences at crowd events required him to visit these events of interest and distribute paper surveys for data collection. Experimental lab research also poses disadvantages, including time and recruitment issues:

"it's difficult to get people into the lab in the first place, to recruit them, it takes a lot of time to organise 'cause you can only do 5 or 6 a day at most, and then once they've done that study they can't really do another one"

Regarding this issue, P2 discussed the benefits of using mobile methods, particularly in gathering data in difficult settings, and eliminating the need to manually transcribe data from paper surveys:

"you can collect data in the field far more easily, collect data in a variety of settings. You don't need a desk to sit and write, you don't need paperwork to collect"

Jeeves could enable self-monitoring

P2 suggested that with intellectually disabled participants, *Jeeves* could be used as a monitoring tool to self-regulate and record dietary habits. The possibility for participants to view and track their data over a period of time, and to receive automatically prompted feedback from an app, could support them to independently manage their health and well-being.

"People with intellectual disabilities are more likely to have various forms of epilepsy because of cognitive damage. So again it would be again your parallel with diabetes, they would be more able to monitor the frequency of seizures and...that would give good feedback for consultants and others"

P2 additionally described how mobile phone use in intellectually disabled populations has dramatically increased in recent years, and that such technology is seen as an asset to their independence.

TABLE III
INITIAL REQUIREMENTS FOR RESEARCHERS

Requirement	Feature Required	Possible Solutions
#1: Researcher Collaboration	Cultures of Participation	Shared editing and semantic annotation
#2: Peer-oriented support	Cultures of Participation	Library of example specs, community forum
#3: Low barrier to entry	Software Shaping Workshops	Domain-specific visual programming paradigm
#4: High ceiling of capabilities	Software Shaping Workshops	Workshops that evolve through feature requests

Jeeves could support sensitive participants

Researchers supported the potential for delivery of feedback to participants, both automated and person-to-person. For example, P2 described working with intellectually disabled individuals, who have a carer or guardian whom they contact for support. Incorporating a means for direct communication between participants and carers was thus considered useful in practical applications of *Jeeves*:

"...if you'd something like this with a button that alerted carers, that's better, that's less intrusive than walking about with a band...for self-monitoring, bullying purposes you've got some sort of button where they can communicate with people and say 'I'm not feeling safe'"

As well as enabling direct support from human sources, this could be supplemented with automated support for participants for whom direct researcher contact may not be feasible. P1 endorsed this possibility:

"Wherever we do research where there is a possibility of causing distress we have to take that incredibly seriously...we could automate provision of support to some extent, or at least automate the beginnings of providing support"

Summary

It was promising to observe that the extended features of *Jeeves* were particularly well-received by the interviewed researchers. Each researcher conceived how these features would be conducive to saving time, improving data quality over current methods, and improving participants' overall experience, all of which were considered to be antecedents of adoption. While some advantages were inherent in general smartphone ESM, these were still grounded in the general adoption factors for public-outward EUD.

C. Initial Requirements

The perceived usefulness of a public-outward EUD tool is not only contingent on its theoretical potential benefits, but on its existing functionality that would enable these benefits to be realised. It was noted that these initial requirements were focused less on particular features, but were instead related to concepts in the meta-design approach in Figure 1.

#1: Allow collaboration within/between research groups
In collaboration with peers, or as supervisors to students, the researchers work in teams with varying experience. Two

researchers highlighted a recent "replicability crisis" in science [44], such that research groups developing software may be indirectly developing for future groups to ensure replicability of studies. Both within and between research groups, community support could scaffold ease-of-use [45]. P2 explained how research was typically conducted as part of a team with different specialities:

"You've usually got somebody who's very up on the evidence, very up on the research, but not necessarily technically that competent...then you've got somebody else on the team that says right I know how to do [programming]"

Thus, meta-designers for EUD-ESM must consider not only the usability and quality assurance of developed artifacts (public-*outward* EUD) but also best practices for encouraging collaboration between researchers (public-*inward* EUD) [8].

#2: Scaffold learning through peer-oriented support

Acceptance of technology is contingent on adequate support for learning and applying its features. P1, in transitioning from *SPSS* to *R* statistical software, expressed how such support had enabled her to learn complex functionality. She consults documentation when performing complex tasks, rather than learning how to do these tasks independently:

"By the time I started using it there was that critical mass of people who were developing wikis and stack overflow and this that and the other...I'm not very good at using R but I am good at Googling how to do what I want to do"

As a particular design consideration, an EUD tool should be designed for a spectrum of end-users. At one end are novices, who consult documentation and relevant examples to develop a study fit for their purposes; at the other end are "power users" who explore all features of an interface, and scaffold typical examples that novice end-users could apply.

#3: Allow researchers to perform simple tasks

The various functions in *Jeeves* for tailoring apps to individuals and triggering based on contextual information were perceived as useful to researchers, but could reduce usability by introducing unnecessary complexity to novice users. Researchers valued the idea of pre-created examples with standardised questions, that could then be tailored if necessary. P2 appreciated the ability to tailor based on attributes, but suggested that this functionality could be introduced in time:

"It's a question of...is there a point at which you need to introduce attributes or do you need to have that there from the start? I don't know what the answer to that is. You would always start with survey design or blocks, whichever"

Distinct levels of technical self-efficacy arose within the small sample of researchers, suggesting that while complex functionality should be provided, common tasks should be simple and intuitive to ensure that time is saved initially.

#4: Support a high ceiling of functionality

Complementary to the previous requirement, the desire for complex functionality was also expressed by all researchers. P1 expressed how the "low ceiling" of what she could accomplish with *SPSS* forced her to transition to the more complex *R* software and begin the learning process again:

"SPSS is very easy to pick up, but you reach a point very quickly where what you want to do is beyond the scope of what it really does and then you have to give up and move to R and start at the bottom of the learning curve again"

Moreover, researchers were averse to software that was inflexible to their diverse needs. P3, a researcher with a technical background, resorted to manual development of generic software to accomplish his goal that purpose-built software did not provide:

"In a sense what I'm doing is press-ganging a more generic piece of software into this kind of mechanism. I mean it can be done, but it's sub-optimal in that sense"

From this perspective, it is important that non-programmer researchers are able to request functionality that meets their needs (Figure 1D), otherwise this functionality will be sought in more complex software, or in a professional programmer.

D. Participant requirements

Further considerations were centred on participants' willingness to comply with a study, highlighting how both researchers and participants have their own separate acceptance criteria. Researchers expressed concerns that an app that is difficult to use, or an app that is intrusive regarding the data it collects, will be quickly removed by participants, and suggested requirements for the smartphone app to improve acceptance.

Assure participants of confidentiality

Researchers explained how conducting remote research requires participants to feel assured that their data is being collected with complete confidentiality. The primary means of doing this is by explicitly giving participants assurances throughout a study:

"...making it very clear to them what was involved in terms of sharing of information...it wouldn't be a sophisticated process but it would need to be something that is done very clearly" (P3)

However, assuring participants of sensitive data storage also relies on an app's reliability in the field, such that there is an explicit need for professional appearance and functionality:

"It needs to look professional, and intuitive...if it's really slow, going between pages, then people are gonna give up. People generally have quite a low threshold I think for some of the stuff" (P4)

Support participants with different skills

Populations with mental and physical disabilities are often ideal participants for social psychology research. It was also acknowledged that many participants would have poor literacy, such that other means of communicating information would need to be employed. P2 suggested graphical depictions of instructions and answers to survey questions:

"Who is it suitable for? If you try and make it too text-heavy you're talking about a fairly narrow group of people but if you open it up with emojis and symbols then you've got something that's more user-friendly"

In summary, with respect to public-outward EUD, the acceptability of both the group performing the development

activities, and the group using the developed artifact, must be taken into consideration separately.

E. Facilitating Conditions

As modelled in the UTAUT (but not in the TAM), **facilitating conditions** also arose, which refer to orgnisational and technical constraints that could prevent adoption of EUD in work practices regardless of individuals' formed intentions. Fortunately, university research appeared to impose relatively few organisational barriers, crucial to usage behaviour. Indeed, three of the five researchers expressed an interest in using *Jeeves* immediately, two of whom we are currently assisting in their own projects.

The primary facilitating condition expressed by psychology researchers was the affordability of software, mentioned as a key concern by all five researchers:

"Affordability is obviously a big thing so...one of the reasons we were speaking about Qualtrics because not only does it seem to be the market leader but it's also...we have a university licence for that which is a major, a major issue" (P4)

A second factor relates to software already in use, including *Qualtrics*, and statistical software such as *SPSS* and *R*. Researchers were excited about the new possibilities afforded by *Jeeves*, but to minimise integration time, required it to integrate smoothly with current software:

"Inevitably there's gonna be things it can't do, and so being able to actually integrate smoothly...capacity to have that interoperability, plug-in capability, developing sort of thing would be great" (P3)

V. DISCUSSION

The interviews provided a wealth of information on the work practices of social psychology researchers. Having established the perceived usefulness, ease-of-use, and facilitating conditions of *Jeeves*, our discussion relates these findings to the initial question of adoption factors of EUD in this domain. In particular, we note the recurrent themes of **time** and **quality** in determining acceptance, which we propose are integral to general acceptance of public-outward EUD.

Time appears to be the most critical barrier faced by researchers, thus the time *Jeeves* would ultimately save (*perceived usefulness*) the time it would require to learn and use (*perceived ease-of-use*) and the time constraints of particular research projects (*facilitating conditions*) are determining factors for adoption.

However, the time *Jeeves* would save is contingent on the specific goals of researchers, which are not pre-defined. For example, in our ongoing case studies, time-saving qualities (such as a means to obtain informed consent, or collaboration features) emerged through researchers' direct use, and were not previously considered. This implies that a meta-design implementation, where end-users have a stake in design *during* use, is a necessary factor for sustained adoption. When researchers were presented with *Jeeves*, they were able to articulate their time-saving requirements easily in terms of
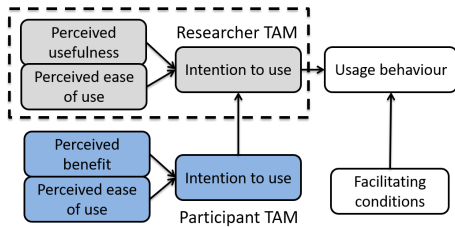
Fig. 5. An informal Technology Acceptance Model of public-outward EUD, requiring benefit and usability to be perceived by all stakeholders

Fig. 6. Interviewed psychology researchers' EUD adoption factors

blocks. Such a representation that allows end-users to communicate their requirements effectively (Figure 1D) is conducive to meta-design, and therefore time-saving features.

Quality is another overarching factor discussed. First, the quality of an app in terms of its functionality is a determining adoption factor (*perceived usefulness*), but particularly in terms of its reliability. A reliable app ensures that constant debugging and participant frustration are minimised (*perceived ease-of-use*), but is also necessary to ensure that apps will not cause harm by malfunctioning (*facilitating conditions*).

Functional quality is critical for adopting new ESM technology. Researchers are already comfortable with using *Qualtrics* software, which fulfills their needs with regards to survey creation. Although software that evolves to the needs of its users is key to ensuring that apps are fit-for-purpose, researchers also have initial requirements that must be satisfied by software which, as P4 expressed, *"let us do that which we couldn't otherwise do"*. While needs vary between end-users, the features derived in Section II, namely: context-sensitivity, participant tailoring, automated feedback, and two-way feedback, were considered desirable by researchers.

A. Public EUD adoption

While researchers have their own model of technology acceptance, this is separate from the acceptance model of their participants. Perceived benefits may overlap, but some are mutually exclusive. Further, initial requirements are also separate, given that researchers and participants interact with two different interfaces. The models are linked, in that researchers will only consider adoption of EUD technology if their participants would be likewise willing to adopt its resultant artifact. Thus, we derived a layered model of technology acceptance, illustrated in Figure 5, to represent public-outward EUD in general.

We also suggest that for adoption of public-outward EUD, it is critical to understand the relationship of domain-experts to their organisation, and to their prospective end-users, prior to engineering. Although we identified two possible applications of ESM, appropriating *Jeeves* (designed as a tool for ESM *research*) into ESM *practice*, requires unique considerations. A summary of researchers' adoption factors, with respect to those that should be high or low, is illustrated in Figure 6.

VI. FUTURE WORK & LIMITATIONS

We further describe two additional areas of future work in research goals described in [37], namely in understanding of stakeholders, and the engineering and evaluation of apps created with *Jeeves*.

A. Understanding - Model investigation

While the socio-technical model of ESM development illustrated in Figure 1 was designed with improved acceptance in mind, we did not directly inquire about researchers' perceptions of specific aspects of the meta-design approach. As part of our ongoing research, we are conducting case studies with researchers to obtain a more in-depth analysis of the utility of these features. Further, we did not attempt to quantify the weight of particular factors in our extended TAM in Figure 5. Future work could probe the true value of these factors in predicting adoption of public EUD in work practices. It is also currently not clear whether this model would generalise to other domains of public-outward EUD.

B. Engineering - Debugging

Application quality and trust are of particular importance when engaging in public-outward EUD. Without knowing how an app will behave prior to its deployment, unexpected functionality issues could critically undermine the utility of ESM. The problem is compounded by the heterogeneity of modern smartphones. Given the privacy concerns surrounding personal data, and the sensitivity of participant groups, researchers must have absolute trust in EUD if they are to adopt it in practice. A suitable testing and debugging framework for researchers could alleviate these issues.

VII. CONCLUSION

The introduction of EUD into professional work practices presents challenges beyond ease-of-use. While the gap between users' software requirements and their programming capabilities appears suitable to bridge with an EUD tool, it is important to ask why the gap exists, and how we as computer scientists are best placed to fill it. Qualitative research with potential end-users can inform us of the likelihood of an EUD tool's success, and indeed feedback from interviews has been invaluable in disrupting our assumptions of what ESM apps could or should do for psychology researchers. Our goal in extending *Jeeves* was not to simply append new features, but to develop it into a system that would allow useful features to be proposed and incorporated as required by end-users. Acceptance of EUD technology is a dynamic process that will require continuous feedback from all stakeholders.

REFERENCES

[1] R. Larson and M. Csikszentmihalyi, "The experience sampling method," in *Flow and the foundations of positive psychology*. Springer, 2014, pp. 21–34.

[2] S. Shiffman, A. A. Stone, and M. R. Hufford, "Ecological momentary assessment," *Annual Review of Clinical Psychology*, vol. 4, pp. 1–32, 2008.

[3] B. Lenaert, M. Colombi, C. van Heugten, S. Rasquin, Z. Kasanova, and R. Ponds, "Exploring the feasibility and usability of the experience sampling method to examine the daily lives of patients with acquired brain injury," *Neuropsychological Rehabilitation*, pp. 1–13, 2017.

[4] V. Pejovic, N. Lathia, C. Mascolo, and M. Musolesi, "Mobile-based experience sampling for behaviour research," in *Emotions and Personality in Personalized Services*. Springer, 2016, pp. 141–161.

[5] N. V. Berkel, D. Ferreira, and V. Kostakos, "The experience sampling method on mobile devices," *ACM Computing Surveys (CSUR)*, vol. 50, no. 6, p. 93, 2017.

[6] H. Lieberman, F. Paternò, M. Klann, and V. Wulf, "End-user development: an emerging paradigm," in *End User Development*. Springer, 2006, pp. 1–8.

[7] D. Rough and A. Quigley, "Jeeves-a visual programming environment for mobile experience sampling," in *Visual Languages and Human-Centric Computing (VL/HCC), 2015 IEEE Symposium on*. IEEE, 2015, pp. 121–129.

[8] F. Cabitza, D. Fogli, and A. Piccinno, ""Each to his own": distinguishing activities, roles and artifacts in EUD practices," in *Smart Organizations and Smart Artifacts*. Springer, 2014, pp. 193–205.

[9] A. Mørch, "Three levels of end-user tailoring: customization, integration, and extension," pp. 51–76, Nov 1997.

[10] F. Paternò and V. Wulf, *New Perspectives in End-User Development*. Springer, 2017.

[11] L. Simons, A. Z. Valentine, C. J. Falconer, M. Groom, D. Daley, M. P. Craven, Z. Young, C. Hall, and C. Hollis, "Developing mhealth remote monitoring technology for attention deficit hyperactivity disorder: a qualitative study eliciting user priorities and needs," *JMIR mHealth and uHealth*, vol. 4, no. 1, 2016.

[12] J. Os, S. Verhagen, A. Marsman, F. Peeters, M. Bak, M. Marcelis, M. Drukker, U. Reininghaus, N. Jacobs, T. Lataster *et al.*, "The experience sampling method as an mhealth tool to support self-monitoring, self-insight, and personalized health care in clinical practice," *Depression and anxiety*, vol. 34, no. 6, pp. 481–493, 2017.

[13] P. Klasnja and W. Pratt, "Healthcare in the pocket: mapping the space of mobile-phone health interventions," *Journal of Biomedical Informatics*, vol. 45, no. 1, pp. 184–198, 2012.

[14] V. Pejovic and M. Musolesi, "InterruptMe: designing intelligent prompting mechanisms for pervasive applications," in *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, 2014, pp. 897–908.

[15] M. N. Burns, M. Begale, J. Duffecy, D. Gergle, C. J. Karr, E. Giangrande, and D. C. Mohr, "Harnessing context sensing to develop a mobile intervention for depression," *Journal of medical Internet research*, vol. 13, no. 3, 2011.

[16] N. Ramanathan, D. Swendeman, W. S. Comulada, D. Estrin, and M. J. Rotheram-Borus, "Identifying preferences for mobile health applications for self-monitoring and self-management: focus group findings from HIV-positive persons and young mothers," *International journal of medical informatics*, vol. 82, no. 4, pp. e38–e46, 2013.

[17] J. D. Runyan, T. A. Steenbergh, C. Bainbridge, D. A. Daugherty, L. Oke, and B. N. Fry, "A smartphone ecological momentary assessment/intervention "app" for collecting real-time data and promoting self-awareness," *PLoS One*, vol. 8, no. 8, 2013.

[18] K. E. Heron and J. M. Smyth, "Ecological momentary interventions: incorporating mobile technology into psychosocial and health behaviour treatments," *British Journal of Health Psychology*, vol. 15, no. 1, pp. 1–39, 2010.

[19] M. E. Hilliard, A. Hahn, A. K. Ridge, M. N. Eakin, and K. A. Riekert, "User preferences and design recommendations for an mhealth app to promote cystic fibrosis self-management," *JMIR mHealth and uHealth*, vol. 2, no. 4, 2014.

[20] J. E. Palmier-Claus, I. Myin-Germeys, E. Barkus, L. Bentley, A. Udachina, P. Delespaul, S. W. Lewis, and G. Dunn, "Experience sampling research in individuals with mental illness: reflections and guidance," *Acta Psychiatrica Scandinavica*, vol. 123, no. 1, 2011.

[21] T. Ludwig, J. Dax, V. Pipek, and D. Randall, "Work or leisure? Designing a user-centered approach for researching activity "in the wild"," *Personal and Ubiquitous Computing*, vol. 20, no. 4, pp. 487–515, 2016.

[22] K. Shilton, N. Ramanathan, S. Reddy, V. Samanta, J. Burke, D. Estrin, M. Hansen, and M. Srivastava, "Participatory design of sensing networks: strengths and challenges," in *Proceedings of the Tenth Anniversary Conference on Participatory Design 2008*. Indiana University, 2008, pp. 282–285.

[23] L. Dennison, L. Morrison, G. Conway, and L. Yardley, "Opportunities and challenges for smartphone applications in supporting health behavior change: qualitative study," *Journal of medical Internet research*, vol. 15, no. 4, 2013.

[24] P. Markopoulos, N. Batalas, and A. Timmermans, "On the use of personalization to enhance compliance in experience sampling," in *Proceedings of the European Conference on Cognitive Ergonomics 2015*. ACM, 2015, p. 15.

[25] S. Vhaduri and C. Poellabauer, "Human factors in the design of longitudinal smartphone-based wellness surveys," in *Healthcare Informatics, 2016 IEEE International Conference on*. IEEE, 2016, pp. 156–167.

[26] T. S. Conner, "Experience Sampling and Ecological Momentary Assessment with Mobile Phones," 2015.

[27] L. F. Barrett and D. J. Barrett, "An introduction to computerized experience sampling in psychology," *Social Science Computer Review*, vol. 19, no. 2, pp. 175–185, 2001.

[28] http://www.surveysignal.com, accessed: 3rd April 2018.

[29] http://www.lifedatacorp.com, accessed: 3rd April 2018.

[30] http://xs.movisens.com, accessed: 3rd April 2018.

[31] A. Gaggioli, G. Pioggia, G. Tartarisco, G. Baldus, D. Corda, P. Cipresso, and G. Riva, "PsychLog: A mobile data collection platform for mental health research," *Personal and Ubiquitous Computing*, vol. 17, no. 2, pp. 241–251, Feb 2013.

[32] http://www.ethicadata.com, accessed: 3rd April 2018.

[33] http://www.pacoapp.com, accessed: 3rd April 2018.

[34] D. Ferreira, V. Kostakos, and A. K. Dey, "AWARE: mobile context instrumentation framework," *Frontiers in ICT*, vol. 2, p. 6, 2015.

[35] H. Tangmunarunkit, C.-K. Hsieh, B. Longstaff, S. Nolen, J. Jenkins, C. Ketcham, J. Selsky, F. Alquaddoomi, D. George, J. Kang *et al.*, "Ohmage: a general and extensible end-to-end participatory sensing platform," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 6, no. 3, p. 38, 2015.

[36] H. Xiong, Y. Huang, L. E. Barnes, and M. S. Gerber, "Sensus: a cross-platform, general-purpose system for mobile crowdsensing in human-subject studies," in *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, 2016, pp. 415–426.

[37] D. Tetteroo and P. Markopoulos, "A review of research methods in end user development," in *International Symposium on End User Development*. Springer, 2015, pp. 58–75.

[38] S. Greenberg and B. Buxton, "Usability evaluation considered harmful (some of the time)," in *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 2008, pp. 111–120.

[39] A. Namoun, A. Daskalopoulou, N. Mehandjiev, and Z. Xun, "Exploring mobile end user development: existing use and design factors," *IEEE Transactions on Software Engineering*, vol. 42, no. 10, 2016.

[40] D. Tetteroo, P. Vreugdenhil, I. Grisel, M. Michielsen, E. Kuppens, D. Vanmulken, and P. Markopoulos, "Lessons learnt from deploying an end-user development platform for physical rehabilitation," in *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. ACM, 2015, pp. 4133–4142.

[41] V. Venkatesh and F. D. Davis, "A theoretical extension of the technology acceptance model: four longitudinal field studies," *Management science*, vol. 46, no. 2, pp. 186–204, 2000.

[42] V. Venkatesh, M. G. Morris, G. B. Davis, and F. D. Davis, "User acceptance of information technology: toward a unified view," *MIS quarterly*, pp. 425–478, 2003.

[43] F. Turbak, D. Wolber, and P. Medlock-Walton, "The design of naming features in App Inventor 2," in *Visual Languages and Human-Centric Computing (VL/HCC), 2014 IEEE Symposium on*. IEEE, 2014, pp. 129–132.

[44] M. Baker, "1,500 scientists lift the lid on reproducibility," *Nature News*, vol. 533, no. 7604, p. 452, 2016.

[45] M. Spahn, C. Dörner, and V. Wulf, "End user development: approaches towards a flexible software design." in *ECIS*, 2008, pp. 303–314.