

Eliciting reactive and reflective feedback for a social communication tool: a multi-session approach

Hilary Smith, Geraldine Fitzpatrick
Interact Lab, Dept. of Informatics
University of Sussex
UK

[hilarys, geraldine]@sussex.ac.uk

Yvonne Rogers
School of Informatics
Indiana University
USA

yrogers@indiana.edu

ABSTRACT

Gaining feedback from users early in the design of a richly featured, novel social system poses unique challenges. We report on our multi-session, in-context approach to get users to envision how they would use an early prototype in everyday life, combined with projections of how their friends would use it. The prototype is a novel social communication management tool and we required users develop a deep understanding of the complete system over time. Findings from data collected across four sessions show that using personalised task scenarios and giving users longer exposure to an early interactive prototype, combined with peer-to-peer discussion, enables participants to move beyond initial reactions to develop more reflective opinions. Participants were able to overcome first impressions and learning effects, develop deeper understanding of new conceptual models underpinning the system, integrate their understanding of piecemeal components and reflect on own use and use by others in deeper ways.

General Terms

Design, Human Factors.

Keywords

CMC, contact management, multi-session user evaluation, prototype, qualitative user study, social network visualisation.

1. INTRODUCTION

An integral part of a user-centred design approach for interactive systems is to carry out user testing early on in the design process “before the development team gets so enamoured with a concept that they won’t discard it” [13, p 1099]. The findings feed into a redesign and this process is repeated iteratively (see [8, 21]). As stressed by Tognazzini

(www.asktog.com) “If you don’t have user testing as an integral part of your design process you are going to throw buckets of money down the drain.” Many methods are available for doing user testing, the most common being measuring performance, asking users their views and opinions and observing people’s reactions.

Typically, a number of users are asked to evaluate a prototype in a one-off testing session and the findings are collated and fed back into the design process. Many studies have demonstrated the effectiveness of this approach for eliciting ‘reactive’ feedback. By this we mean users’ immediate responses and opinions when initially presented with a prototype. This is most useful for assessing a limited set of functions, features or specific form factors. The utility of this approach, however, is less obvious when needing to get feedback on more complex prototypes, where there is a range of aspects that one is interested in evaluating and where the totality of the system concept cannot be presented nor understood in a single session. In such settings, it is often the case that the user is required to imagine how a partially functioning prototype would be used in a given context [e.g. 2]. To obtain a deeper level of understanding of how the prototype will actually be used in their everyday or working life, participants need to engage in a deeper level of reflection and projection. To achieve this, we argue that it is necessary to do more than a one-off user test.

Our research is concerned with how to ‘drill down’ when evaluating a complex partially functioning prototype that requires the users to envision how it might be used both by themselves and also others, which in turn can affect the way they will use it themselves. The particular prototype that we were interested in evaluating is a social network communication service, called in-touch, that we are developing to enable people within a social network to manage their multiple communications with each other more effectively.

The service is designed as an integrated front-end tool to enable users to manage, organize, monitor and archive messages from a variety of communication technologies in one place. It offers a rich set of tools, reflecting functionality that might be found in all of the constituent communication technologies. Hence it can seem initially familiar but is in fact highly novel in the way these technologies are brought together. The service is also meant to be used within a group of friends rather than an individual (although each person will have their version). To

support friends together, it provides contextual features that each can be, or needs to be, aware of others' use. As such it has a number of shared features requiring each user to take another's (e.g. their friends) perspective, imagining what the friend would think about how they were using and setting up their social contact management tool. For example, a projection of self is made through filling out a profile of personal information such as likes and dislikes; current mood can be displayed for others to gauge how one is feeling.

Hence, in addition to getting the initial reactions from potential users to the prototype, we faced two challenges. One was to introduce a rich novel service in a way that participants could give feedback on the overall concept and functionality. We also wanted to evaluate the prototype from a more reflective perspective that took into account the social context. In particular we were interested in the extent to which we could tap into people's ability to understand that the desires and intentions of others are different from one's own and how this understanding affects one's own behaviour when doing something (in this case using the social contact tool). Obtaining such a level of reflection, however, requires the participants attain a level of understanding of the intentions of the prototype so that they can start thinking about how they would use the tool in the context of how their friends would also use it.

The contribution of this paper is the presentation of a multi-session in-context evaluation approach for novel social interactive systems. We devised this approach to tap into a user's understanding of how the prototype could be used in this form of imaginary social context. To this end, the particular methods we used were personalised task scenarios, semi-structured interviewing and a novel form of peer-to-peer reflection. The paper describes our methodology and the rationale behind it. It then introduces the in-touch prototype that was developed for supporting social contact management. The findings from the evaluation study are then presented, focussing in particular on how these findings emerged across the sessions. The paper ends by discussing the benefits of carrying out a multi-session evaluation using different methods in terms of how it can enable a better understanding of the social context of use of a prototype – given the obvious additional costs of doing so.

2. MOVING TO A MULTI-SESSION APPROACH

A single evaluation session with a prototype typically elicits participants' initial *reactions* to what they see and interact with. For example, when evaluating high fidelity prototypes for the first time, it is well known that users tend to focus on the details of the interface (e.g. colour of icons, size of font) rather than the overall structure. Where there is a rich set of potential functionality to be tested, various adaptations of a single-session approach are often used. For example, by virtue of being limited in time, designers can make selective decisions about what aspects or features to test depending on the open questions at hand and so structure iterative sessions to test subsets of functionality with all participants. When it is important to test more features than can be covered in a single session, another approach is to divide functionality across participants, getting

coverage of the whole system but where no one participant has an understanding of the whole system.

To obtain a more *reflective* understanding of the potential use of a prototype, especially in terms of its perceived use by self and others, requires getting participants to move beyond surface level 'look and feel' concerns. It also requires that participants have been exposed to the system as a whole rather than to piecemeal or selected features. We suggest that this level of feedback is more likely to emerge with repeated exposure to a 'full service' prototype over a period of time and through using different techniques of knowledge elicitation. A key question this raises, however, is which methods to use and how to get the respondents to move between reactive and reflective modes when assessing a prototype.

A well-known approach is to use a triangulation of methods [17]. Here, the aim is to replicate the findings from one study in another by using different methods. The rationale for this approach is to increase the generality and construct validity of the results [e.g. 1]. Mackay [14] has also used the triangulation approach to examine different types of data within the same study. Her argument was that, in so doing, a deeper understanding of the problem space can ensue. In a similar vein, we propose that using multiple methods within the same study can result in a better understanding of how the prototype will be used in a social context. However, we also propose that in addition to using a triangulation of methods it can be beneficial to do so over a period of time, involving multiple sessions. In the case of in-touch, time serves two functions: one is that the very nature of the interactions and communications that the tool seeks to support happen over time rather than within a single event; the other is that time allows participants to be exposed to the full feature set of the service, to revisit their first reactions and to be able to develop a more sophisticated understanding of how they and others would use the tool.

So, what methods are most effective for facilitating reflection especially over a period of time? Previously, interpretive methods have been used that give the user full control when evaluating a product in terms of what parts are explored [e.g. 3, 7, 15]. This usually requires a fully functioning system in order for users to have a realistic idea of system behaviour, to allow the normal task of communication to continue and to feel engaged in the application's evaluation.

Another approach for obtaining a deeper level of understanding is the use of diary studies. One of the main benefits of this technique is to show how events develop and issues arise over time. For example, Grinter and Eldridge [9] analysed teenage text messaging behaviour by asking participants to log their incoming and outgoing text messages over a seven-day period. Palen and Salzman [19] have also used a form of voice-mail diary to collect qualitative data on mobile telephone usage. Rodden and Wood [22] used interview and questionnaire techniques at the beginning and end of a 6-month study of a digital photograph storage application. Their multi-session methodology offered an opportunity for questions in the second interview to be refined and placed within context of answers in the first session.

Again, all of these approaches assume the existence of a fully functioning system. In the case of in-touch, this raises a chicken-and-egg problem. The development effort to produce a

functioning version that could be left with people to use would have been significant given that it integrates so many communication tools. To determine whether this effort would be worthwhile in the first place, user feedback is needed but single structured sessions would not have given exposure to enough of the system to elicit this level of feedback. Even if development was feasible, it is not clear that people would explore all of the possible features because it would be such a novel service.

We developed a high-fidelity prototype, avoiding high development effort, but sufficient to give users exposure to all proposed features. The methods we chose for our evaluation capture elements of more structured single session user evaluation of early prototypes in the use of *scenario-based tasks*. This enabled participants to actively explore the full range of functions in the context of communication.

The methods also capture the longer-term exposure of more open interpretive sessions by taking place over *four sessions* (to be described below) and including the person's social and physical context. In particular, we *personalised* the content of the prototype and the scenario tasks where the prototype presented their own friends' names and examples of activities they were interested in. It was assumed that using personalised scenarios in this way would make it easier for them to imagine how they would use the system in a real-life context. We also conducted the sessions in *people's homes* where they had their familiar communication tools around them [6, 11]. This was to encourage participants to think about how they currently communicate with others in a social setting, in addition to them being available for interacting with friends as normal. As important triggers for reflection, these interruptions would be absent from a laboratory study.

Our data collection methods included using a *think-aloud* protocol [4] to make explicit the participants' experience, expectations and feelings while undertaking the tasks and interacting with the prototype itself. To access their wider understanding of the issues surrounding expected use, in-context *semi-structured interviews* (similar to a contextual enquiry approach [12]) were used. Answers here included comparisons with use of their existing communication tools (instant messaging, electronic calendars etc.), how they might use the in-touch prototype, and examples of when features could have been useful in the past. *Peer-to-peer based reflection* enabled participants to reflect on how familiar others would use the service in relation to their own use.

It was decided to conduct four sessions in order to try to access a deeper level of understanding (see Table 1). The first session was used to introduce the study and to conduct a preliminary interview in order to populate the tasks and social network tool with people and information relevant to the participant. The next two sessions were designed to provide an opportunity for the participants to incrementally learn about the various categories of functions provided by the prototype and for the participants to give feedback through talking aloud and answering questions asked by the investigators.

The last session was intended to get participants to reflect about how the service could be used in a social context. To do this, participants were asked to invite a member of their own social group to come along to the final session and to introduce the prototype to them, demonstrating how they would see it being

used. This session was in the control of the primary participant (the participant we recruited) in that there was no structure or script for them to follow. The rationale for this peer-to-peer session was that it would enable us to implicitly see the person's overall opinion of the service, determine their prioritised function set and see what aspects they liked/disliked without having to ask them directly. It would also encourage them to reflect more on the service in the context of this particular relationship.

Table 1. Focus of each evaluation session

Session	Who	Purpose and method used
1	Primary participant and investigator	Introduce study, gather material to personalise tasks and prototype. Interview.
2	Primary participant and investigator	Introduce 1st category of tasks. Think aloud method and semi-structured interview.
3	Primary participant and investigator	Introduce 2nd category of tasks. Think aloud method and semi-structured interview.
4	Primary and secondary participants	Introduce prototype to social group member. Peer to peer discussion and reflection.

In terms of pacing, we chose to use weekly sessions to give the participants time for reflection between sessions, but also for the gap to be short enough for them to remember what they had experienced.

The in-touch prototype that we evaluated using this approach is described in more detail in the following section.

3. THE IN-TOUCH PROTOYPE

Most of us use a multitude of communication tools to keep in touch with family, friends and others. These include email, landline phones, face to face, instant messaging (IM), mobile phone calls and text messaging, chat rooms and more recently voiceover IP when playing multiplayer games [10, 24]. Having so many ways of keeping in touch provides us with much flexibility as to how, when and where we do it. However, the cost of having such choice is that it often entails more remembering, more decision-making and more management overhead. For example, one has to remember which communication tool was used last with someone you need to get back to, which one you are awaiting a reply on and decide which communication method to use for whom and when. In addition, the overheads involved in managing multiple address books and archiving the various communications increases when using multiple methods.

Research in the contact management area has tended to focus on enhancing work-based communications by filtering online communications such as email [e.g., 7, 18]. In contrast, our focus was on understanding what kind of contact management tool would benefit people in conducting and managing their

social communication when using multiple technologies not at work but at their leisure.

To this end we carried out two user studies that investigated people's communication behaviours [5, 23]. Our findings indicated that people can often find it tedious and sometimes overwhelming to manage their social contacts using multiple communication technologies [23]. Young people in particular encounter a number of life changes, including changing schools, college, job, living location and circumstances that negatively impact their ability to manage their communication with others.

To help with this process and to transform it into a more enjoyable experience than a perceived chore, we translated the findings into a set of requirements for a new form of social contact management tool that integrated a number of technologies into a person-centric service. A key requirement was to enable users to organise and keep current a multitude of contact details and messages, across many devices, and to keep track of communication with social network members over time.

The prototype was designed around a self-generated representation of a user's personal social network [see Figure 1] and designed to provide users with five broad areas of functionality. These were:

1. Setting up

The prototype was designed to enable a rapid phase of setting oneself up. Optional functionality in this category includes creating a personal profile and inviting others to join the service.

2. Customisation

A central aspect of the prototype is a personal social network map visualisation that allows rapid access to contact and personal details for each network member, in addition to that person's messages. People can be added to the visualisation with minimal data entry and a photograph or placeholder icon, and groups can be created to manage an expanding social network visualisation.

3. Communication management

The integration of multiple message formats into a single tool allows communication to be managed over multiple technologies from one user interface, e.g. video messaging, email, IM, file sharing and SMS text messaging.

4. Message storage

The social network visualisation was designed to enable users to store and revisit messages based on a person-centric model.

5. Reminders

An integrated social calendar, and the creation of image-based and textual reminders, is combined with individual reminder functionality. Icons representing social network members provide rapid, at-a-glance indication of how much communication has been shared with each member of the social network through a contact graph next to the person's photo.

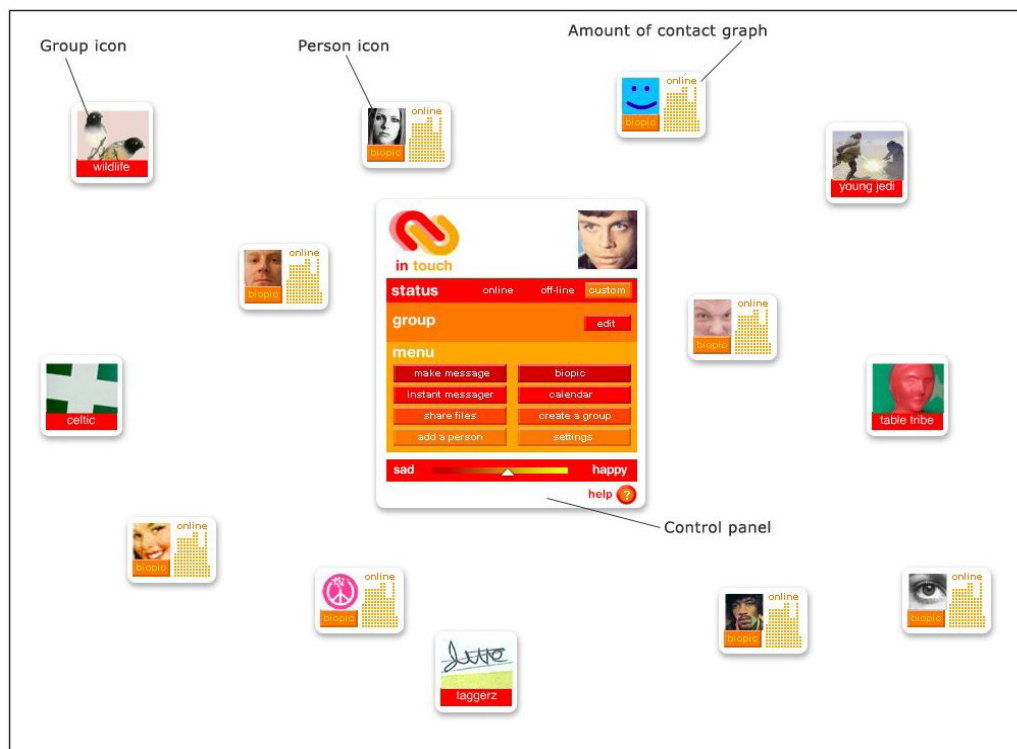


Figure 1. Social network view and control panel

4. EVALUATING THE IN-TOUCH PROTOTYPE

The prototype was initially designed for use with a personal computer or laptop, but envisioned to be used on a variety of devices, e.g. interactive TV, personal digital assistant (PDA), tablet PC. This set-up allowed us to easily evaluate the prototype in people's homes. Thus all sessions took place at the home of the participant, wherever they were most likely to communicate from, e.g., the lounge, family study and bedroom.

4.1 Participants

Ten participants (4 male, 6 female, and aged between 17 and 25 years) were recruited as the *primary participants* from two regional cities in the UK. A second set of participants joined in the final session (*secondary participants*). They were chosen by the primary participants and were aged between 15 and 24 years of age. The gender and age of the secondary participants was not controlled since we wanted our primary participants to introduce their prototype to someone whom they regularly communicated with.

The primary participants reflected a range of social living circumstances. Five of the primary participants were studying full-time, two worked full-time, two were unemployed, and one worked part-time. Six participants lived at home with their parents and four were living independently either in shared accommodation or a young persons' hostel.

4.2 Method

The evaluation sessions were carried out over four weeks, each lasting between 40 and 120 minutes. As previously discussed, session one was designed to introduce the study and to obtain materials so the prototype could be filled with relevant content. Sessions two and three were designed to walk through the personalised scenario tasks, to demonstrate and discuss the five functional areas. Session two showed the start-up related functionality and session three explored features that would be applicable to a more experienced user. In the final session, the participants introduced their personalised prototype system to a friend and discussed how it would be used within their social network.

Participants were paid after each session and received a bonus payment on completion of the fourth session. The secondary participants were also paid.

The last three sessions were either video recorded or audio recorded with synchronised screen-capture. Tapes were subsequently transcribed and analysed. Transcription data from each session was analysed qualitatively. We conducted two passes through the data. One was to draw out the main findings to feed back into the design process. The other was to analyse the data using a grounded theory approach to look for the pattern of reactions and reflections that emerged across sessions. The following section focuses on the latter.

5. EVALUATION OF THE EVALUATION METHODS

The evaluation sessions elicited valuable feedback from the participants on the full range of functionality and on the service as a whole. These findings have been collated into a report and

are being fed back into the next iteration of the prototype. We are not going to detail these findings per se but rather focus on how effective the chosen methods were for eliciting both reactive and reflective feedback and on how different findings emerged and evolved across the multiple sessions. In the following sub-sections, participant comments are referred to by session, and to differentiate between those coming from primary and secondary (i.e. new) participants in the final session, the secondary participants' comments are indicated in brackets.

One of the main findings was the participants' responses did change over time. The final report for the re-design was substantially different to the one that could have been produced from a single session approach. Initial reactions to individual features and functionality were revised as they gained a greater understanding of the service as a whole. This understanding also enabled participants' to reflect more generally on how they would incorporate their disparate tool use into the service and on how they might use the service in their existing relationships. We will return to these themes in more detail when we discuss the impact of multiple sessions. First we review the effectiveness of the other aspects of our approach including personalisation, task-based scenarios and in-context evaluations for encouraging reaction and reflection.

5.1 Personalisation facilitates engagement

Personalisation of the social network presented at the in-touch interface and incorporated into the tasks worked well with this target audience group of 17-25 year olds. It enabled them to readily engage with the relatively complex, unfamiliar system by making it more immediately relevant to their social situation. An unexpected benefit was that it also helped make the tasks fun. They were impressed by accurate scenarios of what a friend had sent them, giggled at unlikely scenarios and offered alternatives. For example, when talking about video messaging one participant said:

"I would never do karaoke though... I'd send my kickboxing moves" Eric, session two.

Personalisation of the prototype facilitated people in becoming aware of how use of the service might change their behaviour. In session four, whilst looking at mutual friends on the social network display, Emma observed that the 'amount of contact' graph was a useful indicator:

"[the graph] would kind of prompt me, and also you have someone's face looking at you. I flick through my phone book and think I should really text such and such and should really phone them and I just don't, but if I had their face looking at me, I'd think 'maybe I should contact you!'" Emma, session four (secondary).

These examples demonstrate that the benefit of incorporating aspects of the participants' real-life into the evaluation is a more comprehensive and personally interesting scenario that can promote engagement and reflection.

5.2 Task-based scenarios ensure system coverage

The personalised content was used to populate a pre-defined set of tasks that were used to structure the participants' interaction with the system, thinking aloud as they did so. These tasks covered all of the major functional areas of the prototype as

previously discussed. If left to their own devices participants might not have explored the prototype as fully because it had such rich functionality and was so novel. Alternatively, system coverage during testing could have been achieved by giving different participants different aspects to respond to. This however, would have made it difficult for any of the participants to gain a deeper understanding of the system.

The approach used here gave all participants the same chance to move beyond reactions to the initial feature set and to explore and reflect on how the broad range of functionality available was integrated into the prototype as a whole, and further, how it might be applied to their own social situation. It also ensured that all primary participants came to the fourth session with similar knowledge about the prototype and were presenting to their friend their chosen aspects from the same background of knowledge. This complete and thorough introduction enabled a deeper reflection that we illustrate in the multiple sessions section below. What it doesn't give us is any strong sense of the relative priorities participants would attach to aspects of the service since we asked them to respond to all aspects equally. The peer-to-peer session was useful in this regard.

5.3 In-context evaluation provides triggers for reflection

Participants enjoyed the novelty of talking about a new communication prototype in their home environment. They were able to carry on monitoring other computer screens for message arrivals, and were available for any contact made by friends or family. There were a number of occasions when sessions were interrupted by incoming communication e.g., text message and mobile phone calls. Rather than view these as being disruptive they were seen as facilitative: helping reinforce the ad-hoc nature of social communication, and on occasion initiating examples of communication behaviour with that person. The following example occurred when a SMS message arrived from Ann, whilst looking at the video messaging functionality:

"...see Ann would really love this, we've tried getting video messaging working between us but her dial-up link and PC are just too slow" Jon, session two.

5.4 Multiple sessions allow for integrated understanding, consolidation and reflection

Complementing the use of task-based scenarios and in-context evaluation, the multiple-session approach enabled the participants to move beyond initial reactions to confirm or revise opinions and to reflect more generally on the service as a whole, how it would fit in with their own communication tools and how they and their friends would use it.

5.4.1 Initial reactive feedback to the prototype

Participants were introduced to the prototype in session two, with some additional functionality introduced in the tasks of session three. Impressions in these early sessions tended to be reactive, concentrating on the look and feel of the user interface and on basic interaction styles. The use of colour to show related people, screen layout, icons, scroll bars and button labels were frequently mentioned:

"The flashing strip would be annoying if I didn't want to read [the new message] now." Sally, session two.

"It's nice seeing everything laid out like that, it is easier than just seeing a list [of people]. It's nicer, I like the look." Julie, session two.

Examples of unexpected behaviour in the prototype were also raised. When typing a message turn in the IM window, Sally said,

"I should be able to type return here rather than having to press send." Sally, session 2.

Emma also found it frustrating when filling in a number of optional fields to add a friend to her social network:

"You can't use the tab through function on this, I would like to be able to tab through." Emma, session 2.

The reasons for such reactive responses can be many and varied; in the early exposure to functionality, especially in a novel system, it is not clear how much of this initial reaction is to do with a learning effect and how much reflects a 'real' opinion of the participant. The multiple sessions provided a way of triangulating these initial responses with later responses since the participant is likely to re-visit many of the same features in different contexts. The following sections illustrate examples of where initial reactions become confirmed opinions while others were revised entirely.

5.4.2 Confirmation of reactions into opinions

Some of these initial reactions became reinforced and consolidated into firm opinions. One aspect that many people remained adamant about was their dislike of the 'loves and hates' section of the personal profile, as indicated by Jon:

"I wouldn't bother to fill in all the love and hates part... My friends wouldn't either." Jon, session two.

"[The profile part] should have been cooler. Loves and hates are a bit cheesy. I wouldn't bother." Jon, session four.

The indicator of a user's 'current mood' provides an option to tell friends how they are feeling on a continuum between sad and happy. Reactions to this feature were mixed, and for some who evaluated it negatively when introduced, these opinions did not change over time:

"I wouldn't use it personally. I quite like people not knowing how happy and sad I am. It's one of the joys of not seeing them face to face." Pete, session two.

When showing the feature in session four to Mary, he said:

"You can set your happy and sad thing, not that I would personally do that," Pete, session four.

Over time it was evident that some reactive comments were revisited and revised as detailed in the following section.

5.4.3 "I've had a think about this": changing initial reactions

Other initial opinions became revised over time as new knowledge was assimilated into their understanding of the prototype and participants were able to reflect more generally on the prototype. This was noticed in two areas in particular: revision of initial conceptual models; and awareness of how use of the prototype would change their requirements.

Understandably, some aspects of the service were familiar, the experience that our participants had gained from other

communication tools, and their associated conceptual models, was brought along to the evaluation of this novel communication tool. As could be predicted, this had a strong effect on initial reactions and interpretations of what the prototype was about.

The use of email was one example. The in-touch method for addressing emails, a function that was familiar to many participants, was initially confusing for some. The method required the user to close the message window before inserting or giving a recipient name or email address. Closing the message window revealed a small envelope icon that is dragged and dropped on to one or more recipients, including whole groups of people. Many people initially tried to drag people icons onto the message window or drag people to the envelope icon but ultimately participants overcame their initial unfamiliarity and started to understand and like the model:

"It's easy once you know how" Sally, session two.

"It's just like posting them a letter" Julie, session two.

For another participant the method was much more problematic:

"That's what I'm used to – writing who before I write the message. I don't think I like that... I don't like having the indicator envelope dotted all round the screen. I want to be able to look at the screen and know exactly who it is going to. Whereas if you dragged the people onto a list you could sit and check them off, you've got a list". Jim, session two.

In the following session, he volunteered that he had been reflecting on the email-addressing method in the intervening time:

"I've had a think about this... the second time around that does make a lot more sense than dragging the person [to the message box]" Jim, session three.

Later on in the session, he further extrapolated this knowledge to other message sending techniques when SMS and video message sending tasks were introduced, confirming that he had overcome the initial learning curve effect and had bought into the new interaction model.

IM provided another example where participants initially relied heavily on their prior experiences with more familiar tools to help them come to terms with a novel tool. For example, they compared the set-up time to their experiences with IM:

"The thing I'm trying to compare it to straight away is MSN messenger to see how it compares with adding people on to that." Sarah, session two.

"Generally that is a lot easier to use than yahoo messenger or MSN. Adding someone you don't know on MSN is really longwinded." Steven, session two.

Steven went on to use his IM tool as a reference point for interpreting many of the features of in-touch in session two, e.g., likening the personal profile to MSN settings and relating the spam filtering to the MSN blocking function. These initial impressions affected how some participants interpreted the tool, and persisted until they had been introduced to a wider set of functionality than is offered by their current IM tools. Over the sessions, comparative references to IM tools significantly decreased. Steven's references to MSN did not happen again after session two as he started to understand in-touch as having

broader functionality than MSN and a different underlying conceptual model.

In addition to the evolution of conceptual models helping to make sense of the system, we found evidence of how people changed their initial opinions as they became more aware of what they could do with the service and how this would change their own requirements.

Alex's reactions to the idea of groups is an example of this. The social network visualisation is the main screen view on in-touch. In session three, the problem of an overcrowded screen regularly came up when participants were introduced to displaying a sub-group of people on its own. Where organising friends' icons into groups had previously been disliked, it now became an alternative to reducing the size of icons and the control panel.

Alex had already identified the potential problem of screen clutter in session two, but he was adamant that he wanted an overall view of everyone, using clustering rather than subgroups to keep *"my Uni friends all bunched together or work friends and family somewhere else."*

Even within the one session however, he started to revisit this idea as he thought about it more and went through other tasks:

"I think it could get too cluttered on the screen like this. How do groups work?" Alex, session two.

In session three, Alex returns to his initial position and reiterates that he does not want to use sub-groups on his social network. However, when he is asked in a later task to address a message to a group, he realises that this would be a useful feature but that in his preferred model, the system would have no way of knowing who was in each of his groups if he just used clustering of people icons.

Further on in session three, Alex is exposed to the group change feature in another task when he has to search through messages and he finally decides he likes it.

"Oooh! I quite like that, the way the groups work, moving between them." Alex, session three.

He ultimately revises his initial position on the basis of repeated experiences with the group idea in different task contexts and decides that groups would actually be useful. Not only this but he reflects further to extrapolate its usefulness to a new situation where he might be using in-touch on a small-screen e.g., PDA. In session four he confidently tells his friend that using sub-groups of people can reduce screen clutter.

5.4.4 Reflecting on their own use of in-touch tools

Initially, participants tended to engage with the individual features of the prototype drawing on existing conceptual models of similar individual tools e.g., IM and email as previously discussed. One of the impacts of giving people comprehensive and prolonged exposure to the service was that they started to understand the service as a whole, reflecting more generally on their own contact management practices and how they might use the system.

In particular, support for the integration of many communication tools into one application was strengthened over time as participants saw the potential benefits of storing all message types in one area to create a more complete person-centric profile of communication over time. When talking about

making calendar reminders in session two, Alex went on to discuss his need for integration with other technologies that hadn't previously been mentioned:

"[Alerts] is something I'm trying to get in to. I've got a PDA and am trying to use that. I forget birthdays and where I am supposed to be. I have it synched in with my PC in work [for backup]... It's like if you lose your mobile phone, I haven't got any of my phone numbers written down. I've been trying to get them into that [PDA] and it's taking forever." Alex, session two.

This started him thinking about his need for in-touch to integrate with other technology. By session four this had developed into a more formalised vision of what he needed.

"I'd want it at work, on PDA, at home. Would have to be on any device and available all the time. I would then use it a lot, I have too many modes and messages." Alex, session four.

In thinking about how they would use the service themselves, many participants even went so far to comment, unprompted, on the business model for offering such a service, suggesting what they would and would not be willing to pay for. This was a clear example of them having integrated the idea of the whole service and projecting themselves as future owners/users.

5.4.5 Reflecting on how they and their friends would use the service

Over the third and fourth sessions it was clear that participants were reflecting not just on their own potential use of the features available, but also imagining how *their use* of the features might affect others in the social network. For example, the in-touch service offers a way to present your current mood to those in your social network by setting a mood indicator on the control panel. When thinking about casual conversation initiation in session four, Jim said:

"You could set it to happy if you'd just got a new job, to get people to ask. I quite like that one. On MSN there's quite a few people I see who are logged in but I don't necessarily talk to them, but maybe if there was something like that I might be encouraged to ask them." Jim, session four.

The novelty of letting others know how you are feeling using the mood indicator also provided some interesting insights into friends' expected use of the feature. Discussions included whether people would tell the truth with the feature, or instead give an indication of what they were happy to talk about. Sally in session four said that some of her friends already do this:

"Maybe people would if they're really happy but not if they're sad, if something really good happened that day. Sometimes when I get back to my desk someone's left a smiley face on MSN and they say 'oh something really good happened today and I wanted to let you know', yes probably for really happy things but not for sad things." Sally, session four.

Julie predicted that the 'sad' setting would be used too frequently to be effective by one friend, leading to reduced communication at times of sadness rather than enquiries of concern:

"Yes, you can imagine Helen who I used to live with, it would be permanent sadness. I don't talk to her, can't cope... You just try and avoid those conversations." Julie, session four.

Other participants said they would *"just leave it in the middle"* (Steven, session two) rather than giving away too many personal clues. In contrast, when thinking about video messaging Sally wanted to show images of herself when away from home and thought it would encourage her parents too:

"I can imagine especially with somebody who's a long way away it would be quite nice... be good to send it home to your parents or something. They might send me my little brother doing stupid things and things like that." Sally, session three.

Further examples of this reflection on how others might also use the service in conjunction with themselves are illustrated in the following section.

5.5 Peer-to-peer introduction as a triangulation and reflection method

The fourth session, where participants were free to introduce the service as they wished to their friend, was particularly useful as a triangulation and reflection method.

In this session the participants used their familiarity with the service, combined with their knowledge of their shared social context, to explain to their friend how the system could be used. This social context reflection is an effective alternative to directly asking the participants about what features they liked/disliked and how they imagined using the service. It also provided a way of triangulating comments made during sessions two and three when performing individual tasks as driven by the task scenarios. Whilst the task scenarios were focussed on presenting the prototype's functionality, this session was concerned with the user's own perspective of the system. In giving our participants free choice about what aspects of the service they introduced to their friend, we had implicit access to what each felt were the salient and important features. Many common themes emerged, both positive and negative, as higher priorities than others: they strongly disliked elements of the personal profile, felt that mood setting would not be very useful, and liked the novel way of sending email and video messages. The final session helped confirm and prioritise findings from previous sessions.

Through conversation with their friend, participants also showed evidence of having engaged in further reflection about the service as a whole. The role of conversation, collaboration and reflection in learning has been well documented [e.g. 16, 20] as a way of understanding what elements are missing from a learning phase. Transferring new knowledge to another situation helps identify gaps in understanding, and we found that putting the in-touch service into the context of a *familiar other's social network* context provided a powerful stimulant for the issues surrounding use of a novel service.

The familiar social context of the final session meant that the friends knew a lot about each others' lives, and how they might each use it differently. There were many examples where participants suggested ways the tool could be used that were not explicitly pointed out in the previous sessions and what the tool might mean for their friend; this is evidence of them understanding, applying and envisioning further use of the system.

In the following example, Alice has thought about her friend's particular needs and what the implications might be for too little space for multiple email addresses:

"You can add different bits like your email, mobile or even... I was thinking for you - you need more than one space for your mobile" Alice, session four.

In one final session, Sally was more imaginative and enthusiastic with applications within her friends' social network than she had been with her own:

"for Lizzie you could have stuff about the boys in [her profile]", "yes, things I have given them or that other people have given them", "yes I suppose you could put in a wish list, what the boys needed, updated photos of the boys so every time you log on you can see them." Sally, session four.

In the peer-to-peer discussions we gained insight into members of the secondary participants' social network beyond their shared friends with the primary participant. Lizzie, a secondary participant, reflected further on ways the service could help her mother:

"My Mum keeps track of which [Christmas] cards she sent to which people 'cos she doesn't send the same one each year. With her personal ones, she keeps a note, describes the picture next to the person's name. She has a history of the past 6 or 7 years. She definitely needs a computer program to monitor that." Lizzie, session 4 (secondary).

6. Conclusion

Gaining user feedback in the early design phase of the in-touch system posed particular challenges. We designed in-touch to provide a person-centric way to integrate communication and contact management needs as had been identified in user research. As such in-touch was familiar in many of the constituent features it offered, yet highly novel in its presentation and its integration of multiple communication tools. It was also a system that only made sense in relation to a person's own social network. It would have been difficult to enable participants to move beyond first reactions and their associations with more usual tools to understand the system as a whole using a single session approach. It also enabled them to understand and reflect more deeply on how it could be used by themselves and their friends.

The in-context multiple session approach we used instead resulted in findings about the service that were ultimately quite different than had we relied on data from a single session (equivalent to session two), both in what was liked and not liked and how people perceived the tool more generally.

The multiple sessions and structured task scenarios for participants to walk through were both highly effective in giving participants the time and exposure needed to understand the system as being more than (seemingly familiar) constituent parts. This enabled people to move beyond initial surface reactions to either confirm or change initial opinions about the features and functionality offered by the system and to understand how the components worked together. Issues that seemed important at the beginning, such as a different interaction model to the one they were used to, were easily overcome with use and time to reflect.

The use of a variety of techniques such as a personalised prototype and task scenarios, in-home testing, and peer-to-peer introduction were highly instrumental in providing the participants with triggers and opportunities for deeper reflection, concurrent with their growing understanding of the service. They thought about how they might use the service as part of their own communication management and also how their friends might use the tool in relation to themselves. The peer-to-peer session was particularly effective in providing us with indirect access to what participants really thought about the system by what they chose to introduce, or not, and how they did it. This provided us with a prioritised set of requirements for the next prototype iteration.

The approach of using a peer-to-peer session can certainly be applied more generally. It does not necessarily need to be part of a multi-part evaluation to derive the value it offers for promoting reflection and engagement of the very social context in which the system would be used. We anticipate that this type of multi-session approach with an interactive prototype is beneficial particularly where systems share similar features with in-touch – a system that is richly functioned, somewhat novel and that sits within a social space of interaction. The trade-off to be decided is the order of effort and time required to personalise prototypes and scenarios, and to run the sessions over weeks, against the order of effort to build an integrated, fully functioning system for more real world and complete system user testing.

7. ACKNOWLEDGEMENTS

This work was funded as part of the PACCIT LINK programme, funded by the ESRC, EPSRC and DTI in the UK. We thank our collaborators BT, Chimera and Victoria Real. We thank Mark Brady and Mia Underwood who were involved in activities leading to the design of in-touch. We thank Debbie Diduca for her help with data collection and transcriptions and the study participants for their time.

8. REFERENCES

- [1] Bailey, R.W., Allan, R.W. and Raiello, P. (1992) Usability testing vs. heuristic evaluation: a head-to-head comparison. In *Proceedings of Human Factors Society 36th Annual Meeting*, Human Factors Society, 409-413.
- [2] Barkhuus, L. and Dey, A. (2003) Location-Based Services for Mobile Telephony: a study of user's privacy concerns. In *Proceedings of INTERACT'03*, 709-712.
- [3] Bellotti, V., Ducheneaut, N., Howard, M. and Smith, I. (2003) Taking email to task: the design and evaluation of a task management centered email tool. In *Proceedings of CHI'03, Human factors in computing systems*, 345-352.
- [4] Boren, M. and Ramey, J. (2000, September). Thinking aloud: Reconciling Theory and Practice. *IEEE Transactions on Professional Communication*, 1-23
- [5] Brady, M. and Smith, H. (2002). User Needs Analysis. InTouch project, COGS, Sussex University, UK

- [6] Browne, H., Bederson, B., Plaisant, C. and Druin, A. (2001). Designing an Interactive Message board as a Technology Probe for family communication. <http://www.cs.umd.edu/local-cgi-bin/hcil/rr.pl?number=2001-20>,
- [7] Farnham, S., Portnoy, W., Turski, A., Cheng, L. and Vronay, D. (2003) Personal Map: Automatically Modeling the User's Online Social Network. In *Proceedings of INTERACT'03*, 567-574.
- [8] Gould, J.D. and Lewis, C.H. (1985). Designing for usability: key principles and what designers think. *Communications of the ACM*, 28(3), 300-311
- [9] Grinter, R. and Eldridge, M. (2001) y do tngrs luv 2 txt msg? In *Proceedings of Seventh European Conference on Computer Supported Cooperative Work*, 219-238.
- [10] Halloran, J., Rogers, Y. and Fitzpatrick, G. (2003) From text to talk: multiplayer games and voiceover IP. In *Proceedings of Level Up: 1st International Digital Games Research Conference*, 130-142.
- [11] Hindus, D., Mainwaring, S., Leduc, N., Hagstrom, A. and Bayley, O. (2001) Casablanca: Designing Social Communication Devices for the Home. In *Proceedings of Human Factors in Computing Systems*, 325-332.
- [12] Holtzblatt, K. and Beyer, H. (1996) Contextual Design: principles and practice. In *Field Methods Casebook for Software Design*. Ed Wixon, D. and Ramey, J., John Wiley and Sons, New York. 301-333
- [13] Jacko, J.A. and Sears, A. (1999) The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies and Emerging Applications. Lawrence Erlbaum Associates,
- [14] Mackay, W.E. (1998). Triangulation within and across HCI disciplines. *Human-Computer Interaction*, 13(3), 310-315
- [15] Makela, A., Giller, V., Tscheligi, M. and Sefelin, R. (2000) Joking, storytelling, artsharing, expressing affection: A field trial of how children and their social network communicate with digital images in leisure time. In *Proceedings of Human Factors in Computing Systems*, 548-555.
- [16] Miyake, N., Masukawa, H., Yuasa, K. and Shirouzu, H. (2002). Intentional Integration Supported by Collaborative Reflection. *Computer Support for Collaborative Learning*, 605-606
- [17] Morse (1994) Designing funded qualitative research. In *Handbook of Qualitative Research*. Ed Denzin, N.K. and Lincoln, Y.S., Sage Publications, Thousand Oaks, CA.
- [18] Nardi, B., Whittaker, S., Isaacs, E., Creech, M., Johnson, J. and Hainsworth, J. (2002). Integrating Communication and Information Through ContactMap. *Communications of the ACM*, 45(89-95
- [19] Palen, L. and Salzman, M. (2002) Voice-mail diary studies for naturalistic data capture under mobile conditions. In *Proceedings of CSCW, Computer Supported Cooperative Work*, ACM Press, 87-95.
- [20] PlaySpace (2001). Learning SMET in digital playscapes: Research on reflection, knowledge acquisition and knowledge transfer in technology-enhanced, playful learning environments. PlaySpace, <http://playspace.concord.org/Documents/PS%20narrative.pdf>
- [21] Preece, J., Rogers, Y. and Sharp, H. (2002) Interaction Design: Beyond human-computer interaction. John Wiley & Sons,
- [22] Rodden, K. and Wood, K. (2003) How do people manage their digital photographs? In *Proceedings of Human Factors in Computing Systems*, 409-416.
- [23] Smith, H., Rogers, Y. and Brady, M. (2003) Managing one's social network: Does age make a difference? In *Proceedings of INTERACT'03*, 551-558.
- [24] Waters, D. (2003). Melting Pot of Online Gaming. BBC news, <http://news.bbc.co.uk/1/hi/technology/2857061.stm>