Are Explanations Always Important? A Study of Deployed, Low-Cost Intelligent Interactive Systems

Andrea Bunt
Computer Science Department
University of Manitoba
bunt@cs.umanitoba.ca

Matthew Lount
Computer Science Department
University of Manitoba
matthew.lount@gmail.com

Catherine Lauzon
Computer Science Department
University of British Columbia
catherine.lauzon@gmail.com

ABSTRACT
Intelligent interactive systems (IIS) have great potential to improve users' experience with technology by tailoring their behaviour and appearance to users' individual needs; however, these systems, with their complex algorithms and dynamic behaviour, can also suffer from a lack of comprehensibility and transparency. We present the results of two studies examining the comprehensibility of, and desire for explanations with deployed, low-cost IIS. The first study, a set of interviews with 21 participants, reveals that i) comprehensibility is not always dependent on explanations, and ii) the perceived cost of viewing explanations tends to outweigh the anticipated benefits. Our second study, a two-week diary study with 14 participants, confirms these findings in the context of daily use, with participants indicating a desire for an explanation in only 7% of diary entries. We discuss the implications of our findings for the design of explanation facilities.

Author Keywords
Explanations, comprehensibility, transparency, qualitative evaluations, diary studies, recommender systems

ACM Classification Keywords
H.5.2 [User Interfaces]: Evaluation/methodology;

INTRODUCTION
When equipped with the right underlying reasoning mechanisms, Intelligent Interactive Systems (IIS) have enormous potential to improve the manner in which users interact with technology, including providing individualized instruction (e.g., [10]), helping users complete tedious tasks efficiently (e.g., [12]), and helping them cope with large information spaces (e.g., [17]). Potential downsides of incorporating more machine intelligence within an application, however, include an extra layer of complexity and diminished predictability, both of which can make it difficult for end users, particularly those with low technical expertise, to form sufficiently accurate *mental models* (i.e., an internal representation of how a system works [16]). In fact, a lack of comprehensibility or transparency is often viewed as a key usability challenge for IIS, or even an inherent disadvantage of these systems (e.g., [9, 18]).

One strategy for improving transparency is to provide users with information on why and how a system generates its intelligent behaviour. A number of studies have shown such explanations to be beneficial (e.g., [4, 7, 8, 11, 15, 25]). Given the various benefits, one might infer that all systems should include an explanation facility, however, explanations are also difficult to design effectively (e.g., [2, 8]), require time and effort on behalf of the user to consume, and may even lead to lower decision quality for some users [5]. In other words, explanations can be beneficial, but at a potential cost at design time and/or during application use. Consequently, it is important to determine when explanation facilities are most needed.

Despite the large body of work on explanations in IIS, there is still limited information on what properties of an IIS and usage scenarios result in users having information needs that cannot be met through application use alone. This is for two primary reasons. First, most work on explanation utility has focused on complex domains and/or systems that assist users in making relatively high-stakes decisions. There are many examples, however, of what we call "everyday" IIS, such as Google (Fig. 1, right) and YouTube (Fig. 1, left), which assist users in making low-cost decisions, and provide assistance in a generally unobtrusive manner. Many of these systems do not include explanation facilities and the need for such facilities remains largely unexplored. Second, most evaluations have been conducted using research prototypes in a laboratory setting. These types of settings make it difficult to i) assess users' information needs when given the opportunity to interact with an IIS over extended periods of time, and ii) explore realistic cost vs. benefit scenarios, where explanation viewing takes time away from real-world tasks.

Through two primarily qualitative studies, an interview study with 21 participants and a diary study with 14 participants, we examine users' information needs with deployed, low-cost IIS that they interact with on a day-to-day basis. These systems are low-cost in that their recommendations have limited financial, time, and/or safety consequences, in comparison to, for example, a decision-support system for medical diagnoses, where lives could be at stake. Our studies reveal that i) despite having no real explanations available, our participants view the IIS that they use in a positive light; ii) most participants have

IUI'12, February 14–17, 2011, Lisbon, Portugal.

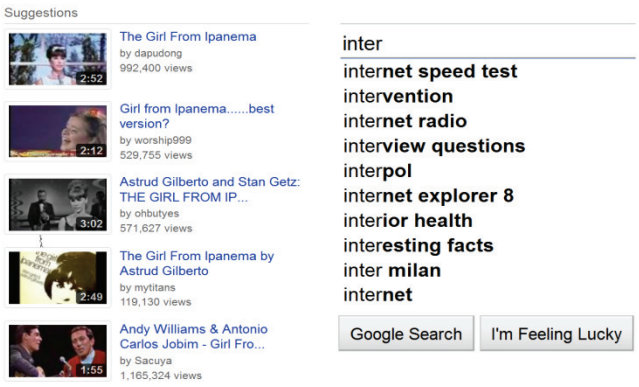

Figure 1. Two low-cost IIS at work: YouTube Suggestions while watching a video and Google Suggest while searching.

formed plausible mental models, although these models are often lacking in detail; and iii) while some users are interested in learning more about how the systems work for general interest or in hopes of improving system behaviour, most feel that the effort to view an explanation would outweigh any potential personal benefit. Together, our findings suggest that explanations are not always necessary to achieve at least operational transparency, and similarly, that not all users need a detailed understanding of how a system works to derive utility from it.

Our studies provide three primary contributions: (1) we describe users' attitudes towards increased transparency in a prevalent class of IIS that has received little attention to date; (2) we present findings based on users' real-world experiences with deployed IIS, used over extended periods of time, to complete real tasks, under real time constraints; and (3) we report qualitative data describing *why* users want or do not want to view explanations.

The remainder of the paper is structured as follows. We first review related work in the area. This is followed by a description of our two studies, including their methods and findings. We then discuss implications for design, study limitations, and promising directions for future work.

RELATED WORK

In this section we review work related to IIS comprehensibility and transparency. We structure our review according to the type of domain studied: expert or knowledge-based systems, ubiquitous systems, complex task assistants, and low-cost recommenders.

Expert or Knowledge-Based Systems

The largest body of work advocating the need for transparency within IIS has been in the domain of decision-support systems and knowledge-based product recommender systems, where explanations are often viewed as a critical component of their success (e.g., [15, 19, 20, 22, 25]). According to Gregor and Benbasat's survey of a number of studies in this domain [7], explanations can improve user performance, helping users make faster and better decisions. Explanations can also improve perceptions of the systems, in particular by increasing levels of trust. The authors suggest, however, that for a user to access an explanation, there must be a specific trigger (e.g., an anomalous result) or a true need to learn how the system works. There is also preliminary evidence that explanations might negatively impact decision quality for some users when the recommender accuracy is low [5].

Ubiquitous Systems

While studies of expert and knowledge-based systems date back to the early 1980s, more recently work has investigated the role of explanations in the domain of ubiquitous computing, or with sensor-based systems (e.g., [3, 13, 14, 24]). Lim and Dey conducted two large-scale Mechanical Turk experiments examining users' information needs with a range of hypothetical ubiquitous applications [13, 14]. Their results indicated that the majority of users wanted more information on how the systems worked, and that when given the experimental task of understanding the system, explanations did improve comprehensibility.

Also in the area of sensor-based systems, and one of the few examples of a relevant field study, Tullio *et al.* [24] conducted a six-week qualitative field study of an interruption management system. Using a set of semi-structured interviews, the study focused primarily on how participants' mental models of the system evolved. The study found that most users had reasonable mental models from the outset, which did not tend to evolve throughout the course of the study, even when provided with explanations.

Complex Task Assistants

In the domain of intelligent task assistants, Glass *et al.* [6] performed a field study evaluating information needs and perceptions of trust of the CALO system, a complex task assistant for a range of office-related tasks. After using the system for a period of 1 day to 2 weeks, participants expressed only a moderate desire for information on why the system behaved a certain way (~3.4 on a 5-point scale), but did indicate transparency as the number one factor influencing their trust in the system. While data gathered in this type of field setting is invaluable, some of the participants in this particular study were highly technically literate, and "several" were CALO developers, which may impact the generalizability of the findings.

Stumpf *et al.* [21] conducted a related study on using explanations to help increase comprehensibility of a machine learning algorithm. Analysis of coded think-aloud statements revealed that certain explanations were preferable to others, but the study did not test whether explanations were needed at all.

Low-Cost Recommenders

The systems described above are either complex, or support high-cost or high-stakes decisions (e.g., large potential impacts on finances, time, or health). We are aware of only two studies focusing on comprehensibility of and

explanations in IIS that support low-cost decisions. Herlocker *et al.* [8] studied the impact of explanations in a movie recommender through two experiments: a laboratory evaluation, and a one-month field study. The laboratory evaluation revealed that certain explanations did significantly improve users' perceived likelihood of viewing a fictitious movie, while others significantly lowered perceptions. The field study revealed that the explanations did not significantly impact users' movie-going behaviour, but that the majority of users wanted the explanations to be included in future versions of the system. A second example can be found in Bunt *et al.*'s study of a system that recommends GUI customizations [2]. While the study revealed that the majority of users appreciated the explanations and wanted them present in the system, there were also users for whom the explanations had little value, and there was no significant impact on user behaviour.

Summarizing Prior Work

In summarizing the above work, we note two main trends. First, the predominant focus has been on complex IIS for which understandability is likely to be most difficult to acquire solely through interactions, and on IIS that support users in making high-cost decisions. The extent to which explanations are needed in lightweight, low-cost IIS has received comparatively much less attention. Second, the majority of evaluations have been laboratory-based, often with the explicit experimental task of understanding how the system works and/or viewing example scenarios instead of interacting with a system. Consequently, their findings do not necessarily generalize to real-world situations, where users both are time-pressured, and have the opportunity to understand and trust a system through long-term use.

STUDY 1: INTERVIEW STUDY

To understand users' desire for explanations in lightweight, low-cost, "everyday" IIS, our first step was to conduct a set of semi-structured interviews with 21 participants (7 male and 14 female).

Study Method

Participants

To understand IIS information needs of an "average" population, we recruited participants who had little formal computer training through a low-level computer usage course and by distributing posters across a university campus (with the exception of the computer science building). The majority of the 21 participants were full-time undergraduate students, with three exceptions: P12 was a civil servant; P13 was a child care worker; and P21 was working in the area of data management. None of the participants had any formal computer training outside of a basic introductory class. Students from the computer usage class received course credit as compensation; the others were given a $10 gift certificate.

Applications Discussed

The focus of our interviews was on eliciting information on IIS that our participants use. If participants couldn't think of any relevant examples, we would then ask if they used a number of applications from a list of popular applications, including: Windows Start Menu, Google Suggest, YouTube recommended videos, Amazon, iTunes Genius, Facebook, and cell phone text messaging. While not an exhaustive list of all deployed IIS, the above list includes a range of system complexities and domains (e.g., GUI customization, low-cost product recommenders, music or video recommenders, and word predictors). Participants were given a second opportunity to describe any other relevant applications after viewing our list.

Explanation facilities within all applications discussed were either extremely shallow or difficult to find without extensive searching, which no participant reported doing. Thus, our findings pertain to systems without a detailed or prominent explanation facility. Table 1 summarizes the capabilities and explanation facilities of all applications discussed by more than two participants.

Study Questions

For each application used, we sought to answer the following:

- To what extent does the participant want to know more about how the system generates its intelligent behaviour?

To further understand factors influencing these information needs, we also sought to answer:

- How does the participant use the application and how does s/he perceive its usefulness?
- How does the participant think the application generates its intelligent behaviour?

Interviews typically lasted from 30 minutes to one hour, and were all recorded. Afterwards, each interview was transcribed in full. We analyzed the transcripts using qualitative analysis techniques from *Contextual Inquiry* [1]. In particular, we created affinity diagrams to identify themes common in the pattern of use and thoughts of participants. In creating the affinity diagrams, participant quotes were clustered based on similarity using a bottom-up, inductive analysis approach. These clusters were then examined for emergent themes. These themes were revised during data interpretation sessions held among the paper authors where we continually revisited the raw data to ensure that the data were consistent with the themes.

Study Findings

To contextualize our findings, we begin by considering whether or not our participants find their "everyday" IIS useful. We then discuss the mental models that our participants have acquired through interacting with the systems. This is followed by a discussion of whether or not our participants desire information on how the IIS work.

Code	Application	Description	Explanation Facilities	Discussed By
YT	YouTube	Recommends videos related to the current video.	Uses the phrase 'because you watched' beside a previously watched video.	P4, P5, P6, P7, P8, P10, P11, P12, P13, P14, P15, P16, P17, P18, P19
CP	Cell Phone	Automatic word corrections or suggestions.	None. Information only available through searching online or through cell phone manual.	P1, P2, P3, P6, P7, P10, P11, P13, P14, P16, P18, P19, P20
AM	Amazon	Recommends products for purchase.	Uses the phrase: 'Customers who bought this item also bought'.	P6, P7, P10, P12, P13, P14, P17
IT	iTunes Genius	Creates playlists from music in a library and suggests music for purchase.	None. Information only available through searching online or through help menus.	P3, P17, P19
FBF	Facebook Friend Finder	Suggests friends for a user's profile on the social media website, Facebook.	Uses the phrases: '# mutual friends' and lists any 'networks' that they share.	P11, P12, P13, P14, P15, P16, P18, P19, P20
FBP	Facebook Page Suggestions	Suggests pages for a user's profile on the social media website, Facebook.	Uses the phrase 'people who like X also like Y'.	P10, P14, P17, P20
GS	Google Suggest	While typing a search into Google, a list of searches is listed that start with the current keystrokes.	None. Information only available through searching online	P4, P5, P7, P8, P9, P10, P12, P13, P14, P15, P16, P17, P18, P19, P20, P21
SM	Start Menu	In Windows, a list of programs that changes according to use.	None. Information only available through searching online or through help menus.	P3, P5, P8, P9, P11, P12, P13, P14, P15, P16, P17, P18, P19, P20

Table 1: Applications that our participants reported using (with >2 discussants) in Study 1.

Perceived Utility

Participants' comments suggest that they generally perceive the low-cost IIS that they use in a positive light, with the primary advantages being time savings and assistance with domain exploration. For example, P18 describes how Google Suggest streamlines his web browsing, and goes so far as to declare speed to be his most pressing concern:

> At some point if you want to go to YouTube, you just hit 'y'; it's the first one... It's always about going quicker and faster when you're browsing the internet. [P18, GS]

In terms of domain exploration, P3 explains how iTunes Genius helps her find available music:

> Actually, it's good because there are a lot of songs that I never knew was [sic] out. I don't listen to the radio as much anymore because of the iPod. [P3, IT]

While most participants find their IIS to be useful and accurate, a number of limitations were discussed. Some feel that the systems lack the sophistication necessary for true personalization. For example, P11 feels that YouTube's suggestions are often redundant, failing to recognize equivalent content between videos:

> If I select one song from one band [...] usually on that recommendation list is just different versions of the same song. [P11, YT]

Others were concerned that recommendations are commercially motivated rather than being personalized:

> Maybe those company [sic] pay money to advertise. [P6, AM]

Thus, participants' comments indicate that they find value in the IIS, but that they aren't uniformly content. None of our participants, however, commented directly on a lack of comprehensibility or transparency as being usability flaws.

Nature of Mental Models

We examine the issue of comprehensibility by describing the mental models that participants have acquired through their interactions with the systems. Our assessments of the accuracy of participants' mental models are based on any explanations found in help menus of the target applications or on-line searching.

Our participants' descriptions of how the systems work suggest a number of mental models (~20%) that are both highly plausible and relatively detailed. For example, in the quotes below, P10 describes how his cell phone makes word suggestions, while P11 talks about YouTube's suggested videos. Both descriptions discuss what sources of data the system considers and P10 also describes how this data is being used:

> It stores a dictionary of words and tries to intelligently guess which word you're going for, based on the grammar and stuff like that – where you're going with the sentence. Words that you use most frequently come up first, and so on. [P10, CP]

> […] different videos have titles, so it's based on text, rather than the content of the video, because I think audio, like the sound and the picture, are much more difficult to capture than the text. [P11, YT]

Other descriptions (~10%) were extremely vague, incorrect, or expressed a great deal of uncertainty. For example, in the quotes that follow, P5 and P15 feel that there is something sophisticated going on, but are not sure exactly what:

> Yeah, what was it... We looked at it in [a computer usage class]. How does it go...? Jeez, I can't remember. Is it the P2P thing? [P5, GS]

> This kind of algorithm that they use, [a] searching algorithm, […] it's like, the strongest algorithm. Probably they can search thousands of data at a time within a second or a millisecond or something. But that's all I know. [P15, GS]

The majority of participants' descriptions (~70%), however, fell somewhere between the above two extremes; plausible explanations, but lacking in detail, and/or uncertain about how different sources of data are combined:

> When watching a video, it will get something similar related to what you are watching. Like if I'm watching something – another video related to that concept, I get more information about what the stuff I was watching [sic]. More information comes up. ... Yeah. The topic. [P8, YT]

> Probably what other people have searched [...] Probably what's on the websites, maybe? I guess sometimes when you searched words before, they usually come up first. [P20, GS].

In the above examples, P8 and P20 have working understandings of what the IIS are trying to do, but not of the details as to how things like "topic" and "similarity" are determined, or exactly what sources of data the system is using to make suggestions. Common sources of confusion were: i) exactly which factors or attributes are considered; ii) how complex factors are assessed; and iii) when multiple factors or attributes are considered, how they are combined.

To summarize, most participants have at least a basic or plausible understanding of how their IIS work, but in most cases, this understanding could be improved through explanations. We next look at whether or not participants are interested in obtaining additional information.

Desire for Additional Information

When initially asked whether they wanted more information on how the IIS they use work, participants' responses were almost evenly divided. Variability in responses tended to relate more to the individual than the application, with certain participants generally interested in more information and others not. As we describe below, reasons behind these initial responses, however, reveal few concrete reasons for wanting additional information, and illustrate some potential challenges for explanation facilities in these types of IIS.

More Knowledge is Never a Bad Thing: Of those instances where participants did want more information, most didn't describe tangible ways in which this knowledge would improve either their interactions with or perceptions of the system. Instead, they indicated that they generally like to know things or thought that the knowledge might help them in future employment. Examples of such sentiments are expressed below:

> In this life you need information. You need to know a lot of things or else you will be left behind. [P8]

> It would be good to know how it works, just like for personal reasons. Just general knowledge. [P2]

> I can tell the kids when they ask me. [P13]

> Because I think there's a very broad market for it in the future, so, maybe I can find a job in it someday. [P11]

Potential for Improved Interactions: The other primary reason for wanting access to an explanation was a hope that viewing one could lead to improved interactions:

> I like to know how things work, and I might find out how to use it better in the future as well. [P10]

> I guess it could help me in my own choice of search terms – to choose the most appropriate thing – if I knew how it worked. [P12]

> Maybe it would help me understand why it doesn't work the greatest. It might help me trick it into recommending better videos. [P10]

In the above quotes, P12 and P10 wish that information on how the system works could help them manipulate the system. P11, on the other hand, hoped that she could use this knowledge to choose to use a product whose reasoning mechanism better suited her needs:

> The next time I choose my cell phone I [could] choose a better one. I will ask which program it's using. [P11]

Two participants also commented on a desire to use the information to maintain their privacy. For example:

> Can I avoid or at least try to avoid things that I don't want known about me to be known? [P21]

Understanding Anomalous Behaviour and Increasing Trust: Two additional justifications for wanting access to more information included helping to understand anomalous behaviour, but for good behaviour as opposed to system errors, and for reasons of trust:

> I would be interested in how it works if they got accurate to the point that it was exactly what I was looking for. [P18]

> I just think it is sort of creepy how it can find things that are just so 'on'. [P17]

> It might affect my opinion of things, and maybe make me trust them more. [P12]

Given findings from prior work, however, we were surprised at the lack of comments relating explanations to trust. P12's comment above was one of only two quotes to directly refer to this issue.

Transparency not Important When it Works: In terms of why participants did not want more information, the most prevalent reason involved only the bottom line mattering:

> I don't really care how things work, just that they do. [P17]

> It wouldn't really bother me if I didn't know, as long as it works. [P15]

> As long as it's working then it's fine. [P1]

In other words, for many participants, detailed transparency is not considered to be important when the system does its job effectively.

Sufficiently Transparent without Explanations: The second major reason for not wanting more information was

participants feeling like they already knew how the system works, or at least knew enough:

> Not that interested. Yeah maybe they will just explain what I said in more complicated terms. [P19]
>
> I cannot explain very clearly, but I think it is like, common sense. [P11]

Examining the mental models of those who indicated they already knew enough revealed a mix in accuracy. Some had quite accurate descriptions, while others had plausible, but incomplete mental models. None, however, gave descriptions that were outright inaccurate. Thus, while some of these participants would likely be able to improve their comprehension by viewing an explanation, none had significant misconceptions that needed repair.

Undesirable Cost vs. Benefit Trade-off: The final reason for not wanting more information involved costs and perceived lack of benefit. A number of participants commented on the time and effort necessary to consume the information. For example, P19 was concerned that any explanation would be overly technical:

> I might find, like, too much technical information about the process. [P19]

For others, there was a lack of apparent benefit:

> I just don't HAVE to know it. [P13]
>
> Are there options for me after I know this information? Like, I know how it works, so I can adjust them, so then it works better for me. Without the second step, it doesn't make any sense. [P11]
>
> [Knowing more] doesn't matter that much. If understanding that would help me type faster, but I think [my] typing speed is already fast. [P16]

In the quotes above, both P11 and P16 indicate that viewing the explanations would have to directly lead to improving their performance with the system either through faster use or making higher quality decisions, with P11 suggesting that he would like the explanation only if he could modify the system to make it better suit his needs.

Discussion

Study 1's findings suggest three primary themes: (1) a high appreciation of these "everyday" IIS despite limited or no explanation facilities available; (2) IIS comprehensibility is not necessarily dependent on explanations; and (3) the cost of explanations is not often perceived to be worth any benefits. We elaborate on these themes below.

High Perceived Value Despite Lack of Explanations

Our participants generally spoke fondly of the IIS they use, with the lack of explanation facilities not negatively impacting general perceptions. Our study does not fully isolate the reason for these positive perceptions, which are likely based on numerous factors, including effective interaction design, commercial success, and the fact that these systems generally work. It could be that explanation facilities are more critical in systems attempting to establish a user base, where accuracy is more difficult to achieve, or with less usable interfaces; however, these results provide encouraging support that if IIS are well designed, users can derive benefit without reading explanations.

Comprehensibility without Explanations

Most users have at least some plausible understanding of how the IIS work. This understanding, while often incomplete, tends to be reasonable enough that it limits many participants' desire for further information. Our study results imply that for at least some IIS, users can acquire workable mental models without the presence of explanations, confirming the generalizability of Tullio *et al.'s* findings [24]. While it is possible that participants could benefit more from IIS with more accurate mental models, a lack of transparency was not perceived to be an inherent limitation despite previous concerns (e.g., [18]). We note, however, that our results do not speak to the perceived utility of explanations in systems where users are not able to build up accurate mental models either because of limited exposure to the system or system complexity.

High Perceived Cost vs. Uncertain Benefits

When using certain IIS as part of their daily lives, our participants indicated that they tend to be more focused on getting their tasks accomplished than on improving their mental models. In fact, a primary attraction of these IIS for our participants is the ability to perform tasks more efficiently than they could autonomously. Consequently, they are wary of explanations that require time and substantial cognitive effort, particularly since they are unsure as to what tangible benefits such additional knowledge might bring.

Gregor and Bensabat [7] indicate that users are not likely to access explanations without a specific trigger, such as anomalous system behaviour. Our results support this notion and raise the question of whether these triggers exist within low-cost IIS. Our participants felt fairly positive about the systems, and ignored the system (or found alternative programs) when it performed poorly. While a couple of participants indicated an interest in seeing explanations for truly amazing behaviour, one might question the likely prevalence of such scenarios.

A second trigger discussed by Gregor and Benbasat [7] is a need to learn how a system works. In a high-cost domain such as medical diagnosis, explanations might be needed to help users explain or justify decisions that they make with the help of the systems to others. The same is not likely to be true of the decision to visit a website or watch a short video. Several participants in our study indicated that there would have to be clearly apparent personal benefit for them to view an explanation. For most of our participants, simply learning about the system was not enough; they wanted it to enhance their use of the system, for example, by making it more accurate, or by increasing the speed gains.

STUDY 2: DIARY STUDY

A potential limitation of Study 1 is that participants' self-reflections might not be representative of their true day-to-day information needs. Thus, to gauge desire for explanations in the context of daily use, we conducted a two-week diary study with 14 participants.

Method

Participants

Participants were recruited through posters distributed around campus. All 14 participants (7 male, 7 female) were either full-time or part-time undergraduate students. None had any formal computer science training beyond an introductory class and none had participated in Study 1. Participants received a $100 gift certificate.

Procedure

Study 2 took place in three stages. In the first stage participants took part in a semi-structured pre-interview, with a format similar to that in Study 1. This was then followed by a 14-day field study, where participants completed one on-line diary entry for each IIS that they used that day. Questions in a diary entry included:

- How useful was the application (scale of 1-7)
- How accurate was the application (scale of 1-7)?
- Did the application behave according to the participant's expectations? Why/why not?
- When interacting with the application, did the participant want more information on how the application works? Why/why not?

Upon completion of the 14-day field portion of the study, participants took part in a post-interview, where further detail on the diary entries was solicited.

Applications

Table 2 lists the applications used by our participants in Study 2 along with the total number of diary entries for each application. Two new applications introduced by participants in this study were: the movie/program recommender in Netflix, and Beatport, an on-line music store with a recommender. Like in Study 1, none of the applications had any explanation facilities to speak of.

Results

In the interest of space, we concentrate our reporting on the diary entries since the pre-interview transcripts revealed attitudes and mental models similar to those in Study 1. Our analysis focuses on the quantitative data present in each diary entry (i.e., the yes/no and Likert-scale questions). We also coded two types of qualitative data. The first was the reasons for why an application matched or did not match expectations. After inspecting the data, we chose a two-level coding scheme: behaved better than expected and behaved worse than expected. The second source was any reasons for wanting more information, which were coded using the themes identified in Study 1 as a starting point.

Application	Discussed By	# Diary Entries
YouTube	P3, P4, P5, P7, P8, P9, P10, P12, P13, P14	35
Cell Phone	P1, P2, P3, P4, P5, P6, P7, P8, P10, P12	51
Amazon	P3, P7, P11	5
Facebook Friend Finder	P1, P3, P5, P7, P8, P12, P13	9
Facebook Page Suggestions	P2, P7	4
Google Suggest	P2, P3, P4, P5, P6, P8, P9, P11, P13, P14	31
Windows Start Menu	P3, P4, P5, P6, P8, P13, P14	15
iTunes Genius	P5, P7, P13, P14	6
Netflix	P2, P3, P4, P5	8
BeatPort	P3	3

Table 2: Applications used in Study 2.

Across all 14 participants, there were a total of 167 diary entries, where each entry represented an entire day's use for a given application. The mean number of entries per person was 11.9 (sd: 4.73). An individual entry reflected anywhere from 1 to greater than 8 uses of the IIS that day, for a combined duration of between 5 minutes and 6 hours. Table 2 shows the breakdown of the entries according to application. Text messaging with cell phones was the most frequently reported application (51 entries), followed by YouTube (35 entries) and Google Suggest (31 entries).

Utility, Accuracy and Matching Expectations

Figure 2 shows the mean ratings for accuracy and usefulness for each application. Means for usefulness ranged from 4.44 to 5.69, whereas means for accuracy were between 4.56 and 6.00. These ratings suggest that participants were generally positive about the systems and that the recommendations tended to be accurate. However, there were also a number of low or neutral ratings (21% of all ratings for usefulness and 23% for accuracy) indicating that there is still room for improvement.

Participants also tended to report that the applications' behaviour matched their expectations (82% or 137/167 entries). Figure 3 (Left) shows the breakdown by application; instances where the application did not match expectations were generally distributed across the applications. Responses for how the application behaved unexpectedly revealed that it was most often because the application was behaving worse than expected (80% or 24/30 instances).

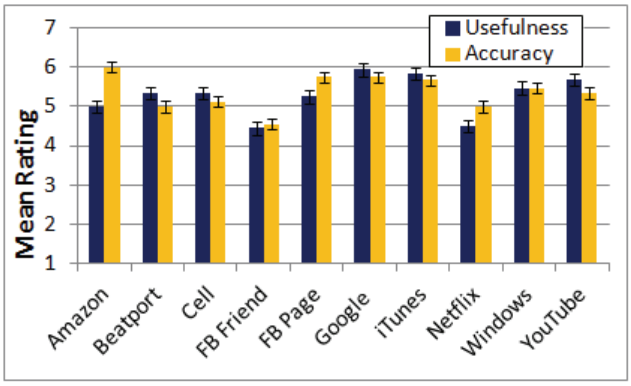

Figure 2. Mean ratings for usefulness and accuracy for each application. Error bars represent the standard error.

Desire for More Information

Participants expressed a desire for more information in only 7% of the diary entries (11 entries). Figure 3 (Right) shows the breakdown by application, indicating that the desire for more information was limited to five applications. These responses also came from only 4 participants; the participant with the largest number of instances of wanting more information (5 of 11) was one of two participants in a more technical field (Electrical Engineering).

We looked at possible triggers for participants' desire for more information. In five of the 11 entries, the participant had reported that the application did not behave as expected, and in all cases, participants comments indicated that the application behaved worse than expected; however, these cases correspond to only 17% (5/30) of the total number of times where participants reported unexpected behaviour.

We also examined accuracy and usefulness ratings to see if poor system behaviour in general was a potential trigger. To do so, we divided the diary entries into two groups: entries where participants wanted more information (*Yes-Entries*) and entries where they did not (*No-Entries*). Usefulness ratings were significantly lower for the *Yes-Entries* according to a 2-tailed independent samples t-test (*Yes-Entries*: 4.6, sd. 2.07; *No-Entries*: 5.51, sd 1.53; p=0.049). For the difference in accuracy ratings, there was also a trend for the ratings for the *Yes-Entries* to be lower (*Yes-Entries*: mean 4.6, sd 1.86; *No-Entries*: mean 5.39, sd 1.48; p = 0.073). Thus, when participants did want more information they were likely to be less satisfied with the application than when they did not. There were many instances, however, where participants were not enthusiastic about the system behaviour, and did not want more information. This was true for 83% (29/35) and 86% (33/38) of usefulness and accuracy ratings, respectively, that were at or below 4 (on the 7-point scale).

Confirming the above analysis, participants' stated reasons in their diary entries for wanting more information were spread fairly even across four main categories: (1) wanting to understand poor or inconsistent behaviour; (2) a desire for improved interactions (even if that meant disabling the feature); (3) wanting to understand atypically good behaviour; and (4) general curiosity.

In the post-interviews, we probed for participants' reasons for *not* wanting more information, which was the case in 93% of diary entries. Responses were reflective of the themes uncovered in Study 1: participants didn't feel like the information would be of personal benefit, didn't want to take any additional time, cared only that the application worked, felt that they already had sufficient understanding gained through interactions with the system, or had built up a high level of trust through interactions. The following quotes illustrate some of the above reactions:

> I kind of figured it out just as I went. […] I didn't have the time to go and look for it, but I guess I learned as I go [sic]. [P7]
>
> I think it's mostly because I don't really feel like it affects my life that much. [P1]
>
> I've been using these things so long that I just generally tend to trust them, or not. [P5]
>
> I don't generally care about how things work as long as it does work. Like if it ain't broke, don't fix it, kind of thing. [P5]

Study 2 Discussion

The results of the diary study reveal that participants very rarely wanted more information on these deployed,

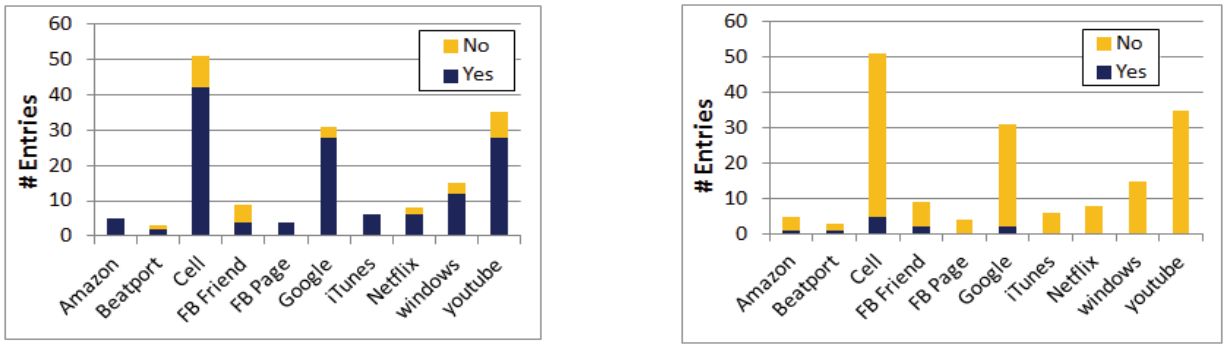

Figure 3. Left: Whether or not the application behaved as expected. Right: Whether or not the participant wanted more information.

lightweight, low-cost systems during the course of their daily use. In fact, in only 7% of the total number of entries did participants express a desire for more information. Those instances tended to have lower usefulness and accuracy ratings than when participants did not want more information; however, in approximately 80% of instances of poor or anomalous system behaviour, participants did not want to know why. Participants' reasons for wanting or not wanting information mirror those in Study 1, providing additional evidence of the validity of those findings.

IMPLICATIONS FOR DESIGN

Our studies help define the boundaries of explanation utility for certain classes of IIS that are used in real-world environments. Without a deeper understanding of when and why explanations are seen as necessary, designers are unable to assess tradeoffs as to where to focus their design/development effort, and how much to promote detailed explanations within their interfaces. While most work on explanations in IIS has advocated their inclusion, our results suggest that explanations are not, in fact, critical in certain systems. Our participants were not overwhelmed by the complexity of the IIS that they use, nor were they confused to the point of not being able to derive benefit.

While our results suggest that explanations are not critical for low-cost, lightweight IIS, there is room to improve the precision of some participants' mental models, and it is possible that improved understanding could lead to enhanced interactions. Our results, however, point to a tension between making explanations highly visible or making them accessible only "on demand". Given our participants' lack of overwhelming interest in obtaining more information, explanations that take attention away from users' primary tasks are not likely to be well perceived. On the other hand, this same reluctance is not likely to result in on-demand explanations being frequently accessed.

In terms of content to include in an explanation facility, the findings from Study 1 point to a number of promising avenues for further study. First, participants doubted that they could personally benefit from viewing an explanation. Therefore, to encourage participants to view explanations, it may be helpful to make any anticipated benefits of viewing an explanation highly visible and within the explanation itself, including information that fosters users' ability to enhance their use of the system. Alternatively, rather than requiring users to use their newly acquired knowledge to adapt their behaviour, an explanation facility could allow users to modify the system's behaviour, or the sources of information that it uses. Second, we found that most users understood the general idea behind the systems, but lacked knowledge on the details, such as how complex concepts (interest, similarity, relevance, etc.) are assessed, or when multiple factors are considered, the manner in which they interact. Thus, if the goal of the explanation facility is to improve mental models beyond what can be derived from interactions alone, a general description is not likely to be sufficient. Finally, participants' desire to invest effort to read and comprehend explanations was low and there was apprehension that the explanations would be too technical. Consequently, it is likely important to keep explanations brief and to use layperson's terms.

LIMITATIONS AND FUTURE WORK

In this section we acknowledge some of the limitations of our studies. First, even the diary study required some degree of self-reflection, since entries were completed once per day as opposed to during system use. Experience sampling would be useful to further validate our findings, by polling users for information needs during system use.

A second limitation of our studies is that participants' reluctance for additional information may have been biased by their pre-conceived notions of what form this information would take. We chose not to introduce any experimental manipulations to gather data on deployed systems used on a regular basis, which happened to have either very minimal or no real explanation facilities, and to avoid a novelty or Hawthorne effect. We feel our results on attitudes towards these deployed "everyday" systems serve as important complements to prior work that has examined different potential formats, but typically in a more controlled and consequently less ecologically valid setting (see [23] for a summary of related work in this area). Viewing different explanation variants, however, might have prompted further discussion and participants might have been more enthusiastic about certain styles of explanations than others.

Finally, our studies indicate that there are systems for which explanations are not crucial, but did not isolate system or user properties that lead to this perception. First, we did not incorporate user expertise into our analysis since we did not screen or test for this factor. Our studies provide initial evidence that some users are more interested in explanations than others, with further evaluations required to understand which user characteristics influence this need. We would also like to explore the impact of the manner in which intelligent support is delivered, focusing on the level of obtrusiveness. For example, the IIS studied here are fairly unobtrusive, whereas IIS with more proactive or interrupting behaviour could require more detailed transparency. Other potential contributing factors to explore are application popularity, commercial success and the level of actual system accuracy.

SUMMARY

In this paper, we described two studies aimed at assessing users' information needs, and consequently the role of explanations in deployed IIS that support users in making low-cost decisions. In doing so, this work is one of the few to examine this class of system and to gather data on non-technical user attitudes towards transparency in IIS based on their real-world experiences.

Our results raise questions as to the importance of, and consequently anticipated usage of, explanation facilities within this type of system. We found that our participants generally perceive these IIS in a positive light despite the lack of meaningful or accessible explanation facilities, and have constructed plausible mental models from their day-to-day interactions with the systems. While some users were interested in accessing more information, the dominant responses were that the applications were sufficiently transparent, or that the cost of viewing an explanation would outweigh the benefit. In combination with prior work on IIS in high-cost domains, our results help define when and why explanations are necessary.

ACKNOWLEDGMENTS

This work was supported by the GRAND Network Centre of Excellence. We would also like to thank Kasia Muldner for her comments on drafts of the paper.

REFERENCES

1. Beyer, H., and Holtzblatt, K. *Contextual Design: Defining Customer-Centered Systems*. Morgan Kaufmann, San Francisco, CA, USA, 1988.
2. Bunt, A., McGrenere, J., and Conati, C. Understanding the utility of rationale a mixed-initiative system for GUI customization. *Proc. UM,* (2007), 147-156.
3. Cheverst, K., Byun, H.E., Fitton, D., Sas, C., Kray, C., and Villar, N. Exploring issues of user model transparency and proactive behaviour in an office environment control system. *User Modeling and User-Adapted Interaction, 15,* 3-4, (2005), 235–273.
4. Dzindolet, M.T., Peterson, S.A., Pomranky, R. A., Pierce, L. G., and Beck, H. P. The role of trust in automation reliance. *Int. J. of Human-Computer Studies 58,* (2003), 697-718.
5. Erlich, K., Kirk, S., Patterson, J. Rasmussen, J., Ross, S., and Gruen, D. Taking advice from intelligent agents: The double-edged sword of explanations. *Proc IUI*, (2011), 125-134.
6. Glass, A., McGuinness, D.L., and Wolverton, M. Toward establishing trust in adaptive agents. *Proc. IUI*, (2008), 227-236.
7. Gregor, S. and Benbasat, I. Explanations from intelligent systems: theoretical foundations and implications for practice. *MIS Quarterly 23*, 4 (1999), 497-530.
8. Herlocker, J.L., Konstan, J.A., and Rield, J. Explaining collaborative filtering recommendations. *Proc. CSCW, (2000)*, 241-250.
9. Jameson, A. Understanding and dealing with usability side effects of intelligent processing, *AI Magazine 30,* 4, *(2009),* 23-40.
10. Koedinger, K.R., Anderson, J.R., Hadley, W.H., and Mark, M. Intelligent tutoring goes to school in the big city. *Int. J. of AIED 8,* (1997), 30–43.
11. Lee, J.D. and See, K.A. Trust in automation: Designing for appropriate reliance. *Human Factors 46*, 1, (2004), 50–80.
12. Leshed, G., Haber, E.M., Matthews, T., and Lau, T. CoScripter: automating & sharing how-to knowledge in the enterprise. *Proc. CHI*, (2008), 1719-1728.
13. Lim, B.Y. and Dey, A.K. Assessing demand for intelligibility in context-aware applications. *Proc. UbiComp*, (2009), 195-204.
14. Lim, B.Y., Dey, A.K., and Avrahami, D. Why and why not explanations improve the intelligibility of context-aware intelligent systems. *Proc. CHI*, (2009), 2119-2128.
15. Pu, P. and Chen, L. Trust building with explanation interfaces. *Proc. IUI*, (2006), 93–100.
16. Norman, D.A. Some observations on mental models. In D. Gentner & A. Stevens (Eds.) *Mental Models*, Hillsdale, NJ: Erlbaum, (1983), 7-15.
17. Resnick, P., Iacovou, N., Sushak, M., Bergstrom, P., and Riedl, J. GroupLens: An open architecture for collaborative filtering of Netnews. *Proc. CSCW*, (1994), 175–186.
18. Shneiderman, B. and Maes, P. Direct manipulation vs. interface agents. *Interactions 4*, 6, (1997), 42–61.
19. Stylianou, A.C., Madey, G.R., and Smith, R.D. Selection criteria for expert systems shells: A sociology technical framework. *CACM*, 35, 10, (1992), 30–48.
20. Suermondt, H.J. and Cooper, G.F. An evaluation of explanations of probabilistic inference. *Proc. Computer Applications in Medical Care*, (1992), 579-585.
21. Stumpf, S., Rajaram, V., Li, L., Wong, W., Burnett, M., Dietterich, T., Sullivan, E., and Herlocker, J. Interacting meaningfully with machine learning systems: three experiments. *Int. J. of Human-Computer Studies 67,* (2009), 639-662.
22. Teach, R.L., and Shortliffe, E.H. An analysis of physician attitudes regarding computer-based clinical consultation systems. *Computers and Biomedical Research 14*, (1981), 542–558.
23. Tintarev, N. and Masthoff, J. Designing and evaluating explanations for recommender systems. In F. Ricci, L. Rokach, B. Shapira & P. Kantor, (Eds.) *Recommender Systems Handbook,* Dordrecht, Netherlands: Springer, (2011), 479-510.
24. Tullio, J., Dey, A.K., Chalecki, J., and Fogarty, J. How it works: A field study of non-technical users interacting with an intelligent systems. *Proc. CHI*, (2007), 31-40.
25. Wang. W. and Benbasat, I. Recommendation agents for electronic commerce: Effects of explanation facilities on trusting beliefs. *J. of Management Information Systems 23*, 4, (2007), 219–249.