

Biases in Human Estimation of Interruptibility: Effects and Implications for Practice

Daniel Avrahami¹, James Fogarty², Scott E. Hudson¹

¹ Human-Computer Interaction Institute
Carnegie Mellon University
Pittsburgh, PA 15213
{nx6,scott.hudson}@cs.cmu.edu

² Computer Science & Engineering
University of Washington
Seattle, WA 98195
jfogarty@cs.washington.edu

ABSTRACT

People have developed a variety of conventions for negotiating face-to-face interruptions. The physical distribution of teams, however, together with the use of computer-mediated communication and awareness systems, fundamentally alters what information is available to a person considering an interruption of a remote collaborator. This paper presents a detailed comparison between self-reports of interruptibility, collected from participants over extended periods in their actual work environment, and estimates of this interruptibility, provided by a second set of participants based on audio and video recordings. Our results identify activities and environmental cues that affect participants' ability to correctly estimate interruptibility. We show, for example, that a closed office door had a significant effect on observers' estimation of interruptibility, but did not have an effect on participants' reports of their own interruptibility. We discuss our findings and their importance for successful design of computer-mediated communication and awareness systems.

Author Keywords

Interruptibility, availability, computer-mediated communication, awareness, ubiquitous computing, context-aware computing.

ACM Classification Keywords

H5.2. Information interfaces and presentation: User Interfaces; H1.2. Models and Principles: User/Machine Systems.

INTRODUCTION AND MOTIVATION

People have developed a variety of conventions that define which behaviors are socially appropriate in different

situations [4]. In office environments, for example, social conventions dictate when it is appropriate for one person to interrupt another. Social conventions also allow the development of an *a priori* expectation of whether or not an interruption is appropriate [17]. For example, a person seeking a colleague's attention is normally able to glance through an open office door, consider a variety of cues, and quickly estimate the colleague's current interruptibility.

This estimation of another's interruptibility, however, is not perfect. On one hand, estimating that a person is more interruptible than they actually are can lead to inappropriately-timed interruptions. On the other hand, estimating that a person is less interruptible than they are can lead to missed communication opportunities. For example, Hudson *et al.* report that managers working with their door closed expressed frustration that people choose not to interrupt them with important information [22].

The use of computer-mediated communication and awareness technology further complicates matters by fundamentally altering what information is available to people. For example, a person calling another's mobile phone typically has no information about the current situation of the person they are calling, resulting in calls at times that are inconvenient, disruptive, or even dangerous for the receiver [2, 30]. This lack of information about a person's current context can have other costs as well. For example, Herbsleb *et al.* report distributed groups considering the lack of a prompt response to email as an indicator that a remote collaborator was irresponsible or did not respect the sender [18]. This phenomena is known as the actor-observer effect, wherein people will often attribute their own behavior to external causes yet attribute the behavior of others to internal causes (such as personality traits), in part because they know more about their own situation [26].

This paper examines discrepancies between a person's reports of their own interruptibility and estimates of that person's interruptibility by other people. Specifically, we examine office environments and ask whether contextual cues, such as whether somebody is talking in an office or whether the door is closed, result in a bias towards

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

CHI 2007, 28 April - 3 May, 2007, San Jose, California, USA.
Copyright 2007 ACM 1-59593-178-3/07/0004...\$5.00.



Figure 1. Sample frames from recordings of a single Reporter participant showing different activities (from left: working on the computer, talking on the phone, and talking to a guest).

overestimating or underestimating the interruptibility of an office worker. A better understanding of such biases could inform the design of computer-mediated communication and awareness systems by providing insight into how people are likely to use different pieces of contextual information. We therefore examine two research questions:

- Which contextual cues affect the error in human estimation of another person’s interruptibility?
- Which of the following is the source of a contextual cue’s effect on the bias in estimation?
 - 1) Failing to consider a cue that is significant
 - 2) Considering a cue that is not significant
 - 3) Overrating the strength of a cue
 - 4) Underrating the strength of a cue
 - 5) Misinterpreting the direction of a cue

To answer these questions, this paper describes new analyses using data from over 600 hours of audio and video recordings collected by Hudson *et al.* and Fogarty *et al.* ([23] and [13]). Fogarty *et al.* learned sensor-based statistical models of office worker interruptibility and compared model accuracy to that of human estimates. In this paper, however, we examine the contexts in which human estimates are unreliable. Our findings show, for example, that the state of an office door is significantly correlated with errors in estimation of interruptibility and that it is likely that people incorrectly treat a closed door as a strong indicator of non-interruptibility. Our findings suggest that people are also prone to overrating the importance of social engagement (such as talking or the presence of guests in an office). The reader should note that, while our analysis allows us to determine a cue’s effect on estimation bias and point to the source of this bias, it does not determine that a cue was directly and consciously used by participants.

This paper makes at least two contributions to the study of human-computer interaction and computer-supported cooperative work. First, we examine a variety of contextual cues to determine which are associated with errors in human estimation of interruptibility. Our results are

valuable for designers of context-aware communication and awareness systems, as understanding these human biases is important when considering what contextual information to include in a system. Second, we raise several questions regarding the tradeoff between abstractions and the direct presentation of sensed information in computer-mediated communication and awareness systems based on ubiquitous and context-aware computing. Though some related questions have been raised in work on media spaces, the nature of ubiquitous and context-aware computing enables new approaches to these problems.

The next section reviews related work. We then present our method, collecting audio and video recordings in the normal working environments of office workers, occasionally asking the office workers to report their own interruptibility, then showing the collected recordings to other people and asking them to estimate the interruptibility of the office workers. This is followed by a discussion of our measures, including a set of contextual variables that we code to indicate what activities are occurring in the collected recordings. Our results section then provides evidence of contextual variables that have a significant effect on estimation error and probes the causes for the bias associated with each of those contextual variables. This is followed by a discussion of our results and a short conclusion.

RELATED WORK

Significant work on computer-mediated communication and awareness systems has been motivated by a desire to encourage the low-overhead communication that naturally occurs when people are located in close physical proximity. For example, Kraut *et al.* showed that the likelihood of collaborative relationships between scientific researchers decreased sharply if a person was not located in the same corridor of a building [27]. The high-level goal of many systems has therefore been to encourage informal, serendipitous communication and thus mitigate the negative effects of physical distance between people.

Early systems generally pursued a rich media approach, using audio/video connections to create the experience of

“being there.” Media spaces were initially explored as switched audio/video connections between four offices and some common areas in two physically remote research labs [6]. The VideoWindow system used a continuous audio/video link to encourage informal interaction between people visiting two lounges on different floors in a research organization [10]. Adler and Henderson reported their experiences with a “direct office share”, a continuous audio/video link between two offices [1]. Cruiser explored pair-wise connections between people working at their desktop, including a cruise functionality analogous to walking down a hallway of open doors [11]. Additional explorations of this rich media approach include, for example, the CAVECAT [29], RAVE [16], and Montage [31] systems.

Later systems shifted away from providing a rich media experience toward exploring the tradeoffs inherent in the sharing of different types of information. Portholes showed that much of the benefit of a media space could be obtained using occasional still photos instead of continuous audio/video feeds [8]. These still photos represented less of a threat to personal privacy and were also less distracting to the people who used it. Hudson and Smith further examined the tradeoff between protecting personal privacy yet still conveying information that is meaningful in an awareness system [24].

The emergence of ubiquitous and context-aware computing has raised some of these same issues in new forums. Instant Messaging programs often use computer activity to indicate that a person is present at their computer. Furthermore, a variety of additional and more complex information can also be sensed and made available to remote users. Tang *et al.* explore Awarenex, an instant messaging client that shares a person’s location and whether they are connected via a desktop computer or a mobile device [32]. Begole *et al.* developed and evaluated Lilsys, which integrated a motion detector and a conversation detector into an instant messaging client to provide a notion of availability for communication that is distinct from presence [5]. Fogarty *et al.* developed and evaluated MyVine, which pursued a similar approach based on a conversation detector, location sensing, electronic calendar information, and recency of computer activity [15]. Similar to work in awareness systems, Lilsys and MyVine explored some of the tradeoffs between providing a variety of contextual cues, their disruption to users, and the potential threat to personal privacy. The use of interruptibility as an abstraction based on sensed context represents one approach to this tradeoff. Work by Dabbish and Kraut confirms that this is an interesting point in the design space, as they show that joint performance in a controlled task is highest when using an abstracted notion of availability (as opposed to providing no awareness information or complete information) [7]. Similarly, results by Avrahami *et al.* showed that, given

contextual information about the receiver, senders would be able and willing to time their messages to accommodate the receiver’s state [2].

In the context of related work, this paper provides insight into how people are likely to misinterpret contextual cues while assessing another person’s interruptibility in the course of using a computer-mediated communication and awareness system. Sensor-based statistical models of human interruptibility [3, 12-14, 20, 21, 25] may be useful in providing cues and abstractions for use in computer-mediated communication and awareness systems, but the underlying contextual cues are still important. For example, systems may be expected to provide an indication of *why* a model thinks a person is interruptible. Because this type of information is so important to ubiquitous and context-aware computer-mediated communication and aware systems, a better understanding is needed of the biases that result from people’s interpretation of contextual cues. The results presented in this paper are an important step towards reaching this understanding.

METHOD

Design

Two groups of participants provided ratings of interruptibility for a large number of real situations. The first group, which we will refer to as *Reporters*, provided self-reports of their own interruptibility while doing their normal work in their normal office environments. The second group, which we will refer to as *Estimators*, viewed audio and video recordings of the Reporters and provided estimates of the interruptibility of the Reporters. Noting that some additional details are available in [13], we present the most relevant aspects of the method here.

Reporters

Participants in the Reporters group provided reports of their own interruptibility while engaged in their normal work in their normal environments. Audio and video recordings were collected using a computer with an audio/video capture-card connected to a small camera and microphone. A grayscale camera with a wide-angle lens was mounted in the office such that both the primary working area and the door were visible. Figure 1 shows several images captured from the recordings. Video was captured at approximately 6 frames-per second, at a resolution of 320x240. Audio was captured at 11 KHz, with 8-bit samples.

To collect measures of how these participants perceive their own interruptibility, self-reports were collected using an experience sampling technique (ESM) [9, 28]. At random intervals (with an average interval of 30 minutes), the recording machine that was collecting the audio and video recordings played a pre-recorded audio prompt asking the participant to “Please rate your current interruptibility.”

Participants responded on a five-point scale, with a response of 1 corresponding to “Highly Interruptible” (they were willing to accept an interruption) and a response of 5 corresponding to “Highly Non-Interruptible” (they were unwilling to accept an interruption). A sign was posted on the recording machine to remind the participant which value corresponded to which end of the scale. Participants were able to respond verbally or by holding up fingers on one hand to the camera. This approach was used, as opposed to using software running on a person’s desktop computer, to maximize compliance by ensuring a participant can respond even if they are away from their desk. Each participant provided self-reports over a period ranging between 3 and 4 weeks.

Estimators

Participants in the Estimator group provided estimates of Reporter interruptibility based on short audio and video clips of the Reporters. Our goal was to simulate a situation similar to that of a visitor stopping by another’s office (in this case, virtually), considering the cues available to them, and deciding whether an interruption is appropriate. Since Reporters were recorded in their private offices, doing real work, we were required to limit the risk of private and sensitive information being exposed. Consequently, each Estimator was first shown still images of the Reporters and asked if they recognized them. Estimators were then shown recordings only of Reporters whom they did not recognize.

A between-subject design was used, with each participant providing interruptibility estimates for 60 clips randomly selected (without replacement) from the audio and video recordings collected in the offices of Reporters. Estimates were provided after each clip. Clips showed a period of either 15 or 30 seconds immediately preceding a prompt for a Reporter’s self-report. The length of the first clip (either 15 or 30 seconds) was randomly assigned, and clips alternated in length hereon. This manipulation was included to determine whether longer clips provided additional useful information to the participants. We should note that in face-to-face situations, people typically have only a few seconds to make an estimate of another’s interruptibility. However, it is not unreasonable to assume that using 15 or 30 seconds may only improve estimates and not hinder them.

Each session started with the experimenter introducing the graphical interface shown in Figure 2. This interface was used by Estimators to provide interruptibility estimates on the same five-point scale used by the Reporters. Participants were informed that they are allowed to watch each clip more than once, and were encouraged to make their estimates as accurately as possible, without concern for speed. Instructions were followed by a training session in which the participant became familiar with the task by providing estimates for six randomly-selected clips.



Figure 2. The interface used by Estimators, who viewed audio and video recordings in order to estimate the interruptibility of an office worker on a five-point scale.

Measures

Contextual Variables

Six paid coders used custom software to code the collected audio and video for occurrences of a large set of potentially relevant contextual cues, including activities and environmental cues. This was based on the theory that environmental indications of task engagement or social engagement will relate to how the Reporters and the Estimators assess the interruptibility of a situation. The set was chosen to include cues that could be reliably identified and coded from the recordings. This custom software presents recordings in 15 second segments and a coder indicates whether each activity or environmental cue occurred during each segment (true vs. false). The Boolean contextual variables coded are:

Reporter activities:

- Whether the reporter is present.
- Whether the reporter is sitting, standing, talking.
- Touch of, or interaction with desk (primary work surface), table (large flat surface, not the primary work surface), file cabinet, food, drink, keyboard, mouse, monitor (gaze at), phone, writing instrument, or papers (including books, newspapers, or loose paper).

Guest activities:

- Whether any guests are present.
- Whether any guests are sitting, standing, talking, or touching the reporter (defined as any physical contact or very close proximity).

Environmental Cues:

- Whether the door is open, is closed (in the case of a cracked door, neither variable was coded).

		Estimated Interruptibility				
		Highly Interruptible		Highly Non-Interruptible		
		1	2	3	4	5
Reported Interruptibility	1	157 6.9%	91 3.9%	40 1.8%	30 1.3%	10 0.4%
	2	79 3.5%	110 4.8%	72 3.2%	36 1.6%	5 0.2%
	3	119 5.2%	202 8.9%	132 5.8%	79 3.5%	45 2.0%
	4	70 3.1%	109 4.8%	115 5.0%	73 3.2%	39 1.7%
	5	66 2.9%	116 5.1%	96 4.2%	142 6.2%	247 10.8%

(a)

Under-Estimation Error

Estimated as being *less* interruptible than reported.

0	-1	-2	-3	-4
	0	-1	-2	-3
		0	-1	-2
			0	-1
				0

(b)

Over-Estimation Error

Estimated as being *more* interruptible than reported.

0				
+1	0			
+2	+1	0		
+3	+2	+1	0	
+4	+3	+2	+1	0

(c)

Figure 4. (a) Confusion matrix presenting Reported interruptibility and Estimated interruptibility, excluding data from two outlier estimators, together with illustrations of (b) under- and (c) over-estimation errors.

- Whether anybody in the office is talking (prior work shows that this is a highly predictive variable [12, 13]).

Agreement among coders was evaluated by re-coding a randomly selected 5% of the recordings, finding 93.4% agreement at a granularity of whether or not each activity occurred within each 15 second interval.

Since a number of the contextual variables are highly correlated with one another, and in order to avoid problems resulting from co-linearity, we made a number of necessary adjustments to our full set of contextual variables; We created a new Boolean variable, *Social Engagement*, defined to be true if any of the following variables is true: Any Talk, Reporter Talk, Guest Talk, Guest Present, Guest Sitting, Guest Standing, and Guest Touch. These other variables were then excluded from further analyses. Similarly, the Keyboard, Mouse, and Monitor variables were combined to create a new variable named *Computer*. Finally, we removed Door Is Open in favor of Door Is Closed, removed Reporter Sitting in favor of Reporter Standing, and removed Table in favor of Desk.

These adjustments yielded the 12 contextual variables used in our remaining analyses. (For the list of variables, see Figure 5.)

Error Measure

Given indications of interruptibility for a situation from both self-reports (Reported) and human estimates (Estimated), our measure of estimation error is the difference score:

$$\text{Estimation Error} = \text{Reported} - \text{Estimated}$$

This error ranges from -4 (corresponding to a Reported = 1 and Estimated = 5) to 4 (with Reported = 5 and Estimated = 1). A negative Estimation Error thus indicates an *Under-Estimation* of the reporter's interruptibility, that is, interpreting a situation as being less interruptible than it

was reported. A positive Estimation Error on the other hand indicates an *Over-Estimation* of the reporter's interruptibility, or interpreting a situation as being more interruptible than it was reported.

It is worth noting that, while the overall accuracy of Estimators is likely to decrease with the use of a fine-grain scale (compared, for example, to a 2-point scale), in this work we are interested in identifying systematic biases in estimation of interruptibility (for which a fine-grain scale is preferred) rather than in overall Estimator accuracy.

Control Measures

Each Estimated interruptibility was assigned, in addition to the corresponding Reported interruptibility, the following control measures: Trial Number (1 to 60), Duration (clip duration: 15 or 30 seconds), ReporterID (the Reporter shown in the clip), and MultipleClipWatch (indicating whether the Estimator chose to watch the clip more than once before giving an estimate).

Participants

Four participants were recruited for the Reporters group, each of whom works in a high-level staff position within a major university, with responsibilities for the day-to-day administration of a large university department and/or graduate program. (The relatively high invasion of privacy resulting from the recording instrumentation, combined with the onerous task of providing self-reports over a month-long period, makes recruiting participants for the Reporters group a difficult task.) The participants had private offices with closable doors, but their responsibilities required them to interact with many different people and they generally did not have full control over their time. These four participants provided a total of 672 self-reports over the course of more than 600 hours of recordings. Of these, 587 were appropriate for use in this study. The others were not used because they revealed private or

potentially sensitive information, or because a technological artifact, such as a small gap in the video shortly before a self-report, might have been distracting to Estimators.

For the Estimators group, we recruited forty participants using a website that advertises experiments within the university. A large majority of Estimators were students at our university or at a neighboring university. Each participant was paid for a session scheduled to last one hour; sessions were not timed but none lasted longer than the scheduled hour. As 40 estimators each provided 60 estimates, each of the 587 self-reports had at least four estimates provided for it, including at least two based on 15 seconds clips and at least two based on 30 seconds clips.

RESULTS

Before describing our results in detail, we first describe the plan of analysis that will follow. We start by presenting two basic and important results. We show that, while a significant relationship exists between Reported and Estimated interruptibility, the two are significantly different from one another. These results confirmed the existence of a bias in estimation of interruptibility, and led to a two-part detailed analysis of the contextual variables. In this analysis, we show contextual variables that have a significant effect on Under-Estimation (biasing Estimators to err in estimating that a situation is less interruptible than reported) and measures that have a significant effect on Over-Estimation (biasing Estimators to err in estimating that a situation is more interruptible than reported). In the final step of the analysis, we use the effect of each contextual cue (or lack thereof) on either Reported or Estimated interruptibility to categorize the source of the cue's effect on bias in estimation. In all the analyses described below, Reported interruptibility, Estimated interruptibility, and Estimation Error were modeled as continuous.

General Effects

We started our analysis by testing whether a correlation exists between self-reports and estimations of interruptibility. We conducted a Mixed Model analysis¹ using the data from our Estimators, with Estimated interruptibility as the dependent variable. Reported interruptibility, Clip Length (15 seconds vs. 30 seconds),

¹ Individual differences had to be accounted for statistically because each participant, whether Reporter or Estimator, provided a number of interruptibility ratings. This was done by including the participant number as a random effect in the model. (Unlike fixed effects, for which the levels of the effect are specifically chosen by the experimenter, random effects can be thought of as those whose levels are randomly chosen out of some larger population and would be different were the experiment to be run again.)

Trial Number, and MultipleClipWatch (0 vs. 1) were modeled as fixed effects. ReporterID (s1, s2, s3, s4) and Participant Number were modeled as random effects².

This analysis showed that Estimated interruptibility is significantly correlated with Reported interruptibility ($F[1,2359] = 463.3, p < .001$). We also found a small significant effect of Trial Number on Estimated Interruptibility ($F[1,269] = 7.41, p < .01$), although the parameter estimate (-0.006) shows that the size of this effect was very small. Still, we see that as the experiment continued, Estimators tended to rate Reporters as more interruptible. One possible explanation is that Estimators tired of the task. Another intriguing possibility is that Estimators began to feel acquainted with the Reporters and felt more comfortable declaring them to be interruptible. Clip Length did not have a significant effect on Estimated interruptibility ($F[1,2354] = 2.00, n.s.$), nor did MultipleClipWatch ($F[1,2366] = 1.48, n.s.$).

Next, we examined the overall effect of Role (as Reporter or Estimator) on interruptibility (either Reported or Estimated). We used a Mixed Model analysis in which Interruptibility was the dependent measure and Role was modeled as a fixed effect. For both types of participants, Participant Number was modeled as a random effect. This analysis showed a significant main effect of Role on Interruptibility ($F[1,34] = 9.27, p < .01$) with participants in the Reporter role reporting that they were more non-interruptible on average ($M = 3.35$) than perceived by Estimators ($M = 2.69$)³.

A closer look at Estimators' data revealed two participants whose estimations were far lower than others and had greater error. In fact, one of these two Estimators never provided an estimate greater than 3 - the middle value on our five-point scale. A Mahalanobis outlier analysis showed the two Estimators to be outliers. We excluded these two outliers from the data and repeated the analysis described above. Role still had a significant effect on Interruptibility ($F[1, 28] = 13.3, p = .001$) with the average interruptibility of Estimators increasing slightly to ($M = 2.76$). While the behavior of the two outliers might be interesting in itself, their estimates (representing 5% of the Estimator data) were excluded from the remaining analyses.

² One could argue that ReporterID should be modeled as a fixed rather than a random effect. Repeating our analyses with ReporterID treated as a fixed effect yielded near-identical results (significance for all other fixed effects remained the same).

³ Because the independent variables were not completely orthogonal, we used Least Squared Means (LS Means) to control for the values of the other independent variables. The means reported throughout this article are LS Means.













Contextual Variable	Effect on Error		Effect on Report Estmt		Bias type	Contextual Variable	Effect on Error		Effect on Report Estmt		Bias type
 Social Engagement	⊖	⊕	Ⓡ	ⓔ	3	 Writing	·	·	Ⓡ	ⓔ	-
 Phone	·	⊕	Ⓡ	ⓔ	3	 Food	·	·	Ⓡ	ⓔ	-
 Desk	·	·	Ⓡ	·	1	 Drink	·	⊕	·	ⓔ	2
 Computer	·	⊕	·	ⓔ	2	 Door Is Closed	⊖	⊕	·	ⓔ	2
 File Cabinet	⊖	·	·	·	?	 Reporter Present	·	·	Ⓡ	·	-
 Papers	·	·	·	·	-	 Reporter Standing	·	⊕	Ⓡ	ⓔ	3

Figure 5. Analysis Summary: ⊖ = A significant effect on Under-Estimation Error, ⊕ = A significant effect on Over-Estimation Error, Ⓡ = Has a significant effect on Reported Interruptibility, ⓔ = Has a significant effect on Estimated Interruptibility.

Bias types: 1 = Failing to consider a cue that is significant, 2 = Considering a cue that is not significant, 3 = Overrating the strength of a cue, 4 = Underrating the strength of a cue, 5 = Misinterpreting the direction of a cue.

Examining Estimation Error

The analyses described so far show that participants are able to make estimates that significantly correlate with self-reports of interruptibility, but also that there is a significant difference between the two.

In order to examine the effect of context variables on errors in estimation, we performed separate analyses of Under- and Over-Estimation errors, each examining a subset of the data (see Figure 4). The first examines which contextual variables correlate with Estimators assessing situations as being *less* interruptible than reported (Figure 4b). The second examines what contextual variables correlate with Estimators assessing situations as being *more* interruptible than reported (Figure 4c). This method allows us to identify and address cases where the same contextual variable introduces both under- and over-estimation (e.g., underestimating interruptibility when a cue is present, while overestimating interruptibility when the cue is absent).

The analyses of Under- and Over-Estimation were done in an identical fashion. We performed a Mixed Model analysis where Estimation Error (once Under- and once Over-Estimation) was the dependent measure and the following measures were modeled as fixed effects: Reported interruptibility, Clip Length (15 seconds vs. 30 seconds), Trial Number, MultipleClipWatch (0 vs. 1), and the 12 contextual variables (see Figure 5). ReporterID (s1, s2, s3, s4) and Estimator Number were treated as random effects.

Finally, in order to determine the cause for a cue's effect on Estimation Error, we examined whether each cue had a significant effect on Reported and on Estimated interruptibility. For example, a contextual variable that does not have an effect on Reported interruptibility but has significant effect on Estimated interruptibility could be characterized as an estimation bias of the form "Considering a cue that is not significant." To do so, we first conducted a Mixed Model analysis using only the Reporters' data, with Reported interruptibility as the

dependent measure. Our 12 contextual variables were modeled as fixed effects. ReporterID (s1, s2, s3, s4) was modeled as a random effect. Next, we used the Estimators' data to conduct a Mixed Model analysis with Estimated interruptibility as the dependent measure. Clip Length (15 seconds vs. 30 seconds), Trial Number, MultipleClipWatch (0 vs. 1), and the 12 contextual variables were modeled as fixed effects. ReporterID (s1, s2, s3, s4) and Estimator participant number were modeled as random effects.

The results of our analyses are summarized in Figure 5. Given the space available, we present detailed results for only the variables found to have an effect on Estimation Error. We provide only a short account of other results.

1. Social Engagement (⊖ ⊕ Ⓡ ⓔ)

We found that Social Engagement had a significant effect on both Under-Estimation (with significantly larger error when the reporter was socially engaged; $M = -.72$ vs. $M = -.56$; $F[1,1133] = 7.38$, $p < .01$) and on Over-Estimation (with significantly larger error when the reporter was *not* socially engaged; $M = 1.09$ vs. $M = .67$; $F[1,1792] = 52.8$, $p < .001$).

Trying to identify the cause for this effect on Estimation Error, we found that Social Engagement played a significant part in both Reported ($F[1,1125] = 73.2$, $p < .001$) and Estimated interruptibility ($F[1,2232] = 127.7$, $p < .001$). In both cases, being socially engaged resulted in a situation being considered less interruptible. We may thus conclude that this is a case of "Overrating the strength of a cue." That is, while social engagement played a significant role in both Reported and Estimated interruptibility, Estimators associate too much importance with this variable, resulting in both under- and over-estimation.

2. Phone (· ⊕ Ⓡ ⓔ)

Phone had a significant effect on Over-Estimation ($F[1,1762] = 75.4$, $p < .001$), with error being higher when the reporter was not using the phone ($M = 1.21$) than when

was ($M = 0.55$). This finding is consistent with the use of a phone representing a reliable indication that a person is not interruptible. Phone did not have a significant effect on Under-Estimation.

As with Social Engagement, Phone had a significant effect on both Reported ($F[1,1157] = 45.23, p < .001$) and Estimated interruptibility ($F[1,2217] = 108.82, p < .001$). Thus, the cause for the effect of Phone on Estimation Error can again be categorized as “*Overrating the strength of a cue.*” This finding could be used, for example, in the design of a system by indicating only that a person is on the phone, but not indicating that a person is *not* on the phone.

3. Reporter Standing (• ⊕ ⊗ ⊕)

Reporter Standing had a significant effect on Over-Estimation ($F[1, 1590] = 10.5, p = .001$), with a greater error when the Reporter was standing ($M = 1.00$ vs. $M = .76$). Standing did not have a significant effect on Under-Estimation.

Looking at each group of participants, we found that a reporter standing had a significant effect for both Reporters ($F[1,1116] = 42.57, p < .001$) and Estimators ($F[1,2139] = 15.12, p < .001$) as an indication that a situation was more interruptible than when the reporter was sitting. Since standing was often part of entering or leaving a conversation or the office, this finding might be consistent with prior work on the link between physical transitions and better times for interruptions [19]. However, this cue’s significant effect on Estimation Error indicates that the cause for the bias is “*Overrating the strength of a cue.*”

4. Computer (• ⊕ • ⊕)

We found that Computer had a significant effect on Over-Estimation, with Estimators being more likely to interpret a situation as more interruptible than reported when the Reporter was using the computer than when not ($M = .98$ vs. $M = .77; F[1,1793] = 11.4, p < .001$). Computer did not have a significant effect on Under-Estimation.

Since we found that Computer had a small significant effect on Estimated interruptibility ($F[1,2238] = 5.87, p \approx .016$), but did not have an effect on Reported interruptibility ($F[1,1160] = 0.02, n.s.$), its bias can be categorized as “*Considering a cue that is not significant.*”

5. Door Is Closed (⊖ ⊕ • ⊕)

The status of the office door had a significant effect on Under-Estimation, with significantly greater error when the reporter’s office door was closed ($M = -.96$ vs. $M = -.32; F[1,1138] = 10.6, p < .001$). In other words, Estimators were likely to misinterpret a closed door and estimate the situation as less interruptible than was reported. Door Is Closed had a marginal effect on Over-Estimation ($F[1, 1795] = 3.65, p \approx .056$).

We found that while the door being closed had no significant relationship with Reported interruptibility ($F[1,1160] = 0.28, n.s.$), it had a significant effect on Estimated interruptibility ($F[1,2243] = 15.0, p < .001$). Thus, we can classify the effect of Door Is Closed as a result of “*Considering a cue that is not significant.*” This is consistent with prior work’s discussion of the common misconception that door status is a reliable indicator of interruptibility [12, 13, 22, 23].

6. Drinking (• ⊕ • ⊕)

Drinking also had a significant effect on Over-Estimation Error, with Estimators assessing Reporters as more interruptible when they were drinking ($M = 1.18$ vs. $M = .57; F[1,1789] = 15.6, p < .001$).

Similar to the state of the door, drinking was significantly correlated with Estimators considering a person to be interruptible ($F[1,2235] = 18.8, p < .001$), though it had no effect on self-reports ($F[1,1160] = 0.05, n.s.$). This variable is thus an example of “*Considering a cue that is not significant.*”

7. File Cabinet (⊖ • • •)

Finally, File Cabinet had a significant effect on Under-Estimation Error, with a smaller error when a Reporter was interacting with their file cabinet than when not ($M = -.44$ vs. $M = -.84; F[1,1133] = 5.90, p < .05$). However, we did not find an effect on Reported nor Estimated interruptibility ($F[1,1161] = 0.12, n.s.$, and $F[1,2239] = 1.47, n.s.$). This effect may be due to some interaction with another variable. Given the limited amount of information available (interaction with a file cabinet occurred in only 3% of the clips), it is difficult to draw conclusions regarding the impact of this variable on Under-Estimation.

Other Variables

Other contextual variables, such as Writing, Desk, Food, and the reporter’s Presence played a significant role in either Reported or Estimated interruptibility, but not in Estimation Error. As expected, Reported Interruptibility had a significant effect on Estimation Error ($p < .001$). We did not see an effect by any of the other control measures (Duration, Trial Number, or MultipleClipWatch).

DISCUSSION

In the previous section we presented results from an in-depth analysis of causes for biases in human estimation of interruptibility, comparing self-reports and estimations based on audio and video clips selected from over 600 hours of recordings collected in actual work environments. Our results identified a number of contextual cues that affected estimation errors and point to their causes. One potential limitation of this work is the fact that Estimators were only able to consider the information available in a

short clip from audio and video recordings. While this is an important limitation if our results are to be interpreted in the context of face-to-face interruptions, we note that the information available to Estimators in this study was similar to information available to users of media space systems, and far richer than information available in most awareness systems. Our results are therefore directly applicable in these contexts, suggesting important considerations for the design of new computer-mediated communication and awareness systems. While we chose not to vary the degree of familiarity between reporters and estimators in this work, further research on the effect of familiarity on estimation bias is needed.

In early audio and video media spaces, large amounts of information about a remote person and their context were shared. This information, however, came at some cost. Personal privacy has been raised as an issue in even the earliest media space work. These rich information streams were also distracting to users. The transition from analog audio and video streams to digital streams allowed systems to process the information, instead of simply mirroring it. This led to systems that shared fewer, but more important bits of information. But this processing has typically been limited to sharing information about presence. Occasional still photos from a low-resolution camera were found useful primarily for determining if a person is present and perhaps whether any additional people are present. Techniques like Hudson and Smith's "shadow view" added more temporal information to the stream [24], but the information conveyed was still primarily related to a person's presence. Our results suggest that providing too much information may not only be a concern for privacy, but may also lead to errors in estimations of interruptibility.

Just as the transition from analog to digital streams allowed new approaches, the emergence of ubiquitous and context-aware computing provides another opportunity to examine what information is shared in computer-mediated communication and awareness systems. Because sensing and recognition systems are now capable of inferring a variety of activities, communication and awareness systems need not be based on a direct presentation of a remote person's activities. Our results suggest that sharing certain contextual cues will likely result in misinterpretations of a person's interruptibility. A new system, informed by our results, could avoid exposing cues that should not be considered in an estimate of interruptibility at a particular time, while enhancing (or moderating) others.

A related question is how systems might manage the fact that only one aspect of a piece of context is prone to misinterpretation. Our results show that being on the phone is a good indication that a person is not interruptible and that it correctly plays a significant role in people's estimates of another person's interruptibility. On the other hand, *not*

being on the phone can be misinterpreted as an indication that a person is more interruptible. How to best take advantage of these types of contextual cues is an interesting question for future research and design work.

CONCLUSION

This paper presented an analysis of the systematic effect of different contextual cues on human estimation of another person's interruptibility, suggesting that careful thought must be given to selection and presentation of contextual cues. The emergence of ubiquitous and context-aware computing now allows new approaches for recognizing and sharing a variety of information about a person's context to be used. Understanding how people are likely to interpret such contextual information is important to designing effective and successful computer-supported communication and awareness systems.

ACKNOWLEDGMENTS

We would like to thank Yaakov Kareev, Darren Gergle, Laura Dabbish, and Joonhwan Lee for valuable insights on this work. This material is based upon work supported by the Defense Advanced Research Projects Agency (DARPA) under Contract No. NBCHD030010, and by the National Science Foundation under grants IIS-0121560 and IIS-0325351.

REFERENCES

1. Adler, A., and Henderson, A. A Room of Our Own: Experiences from a Direct Office Share. In Proc of CHI 1994, ACM Press (1994), 138-144.
2. Avrahami, D., Gergle, D., Hudson, S.E., and Kiesler, S. Improving the Match Between Callers and Receivers: A Study on the Effect of Contextual Information on Cell Phone Interruptions. Behavior and Information Technology (BIT) (In press).
3. Avrahami, D., and Hudson, S.E. Responsiveness in Instant Messaging: Predictive Models Supporting Inter-Personal Communication. In Proc of CHI 2006, ACM Press (2006), 731-740.
4. Barker, R.G. Ecological Psychology. Stanford University Press (1968).
5. Begole, J.B., Matsakis, N.E., and Tang, J.C. Lilsys: Sensing Unavailability. In Proc of CSCW 2004, ACM Press (2004), 511-514.
6. Bly, S.A., Harrison, S.R., and Irwin, S. Media Spaces: Bringing People Together in a Video, Audio, and Computing Environment. Communications of the ACM, 36, 1 (1993), 28-46.
7. Dabbish, L., and Kraut, R.E. Controlling Interruptions: Awareness Displays and Social Motivation for Coordination. In Proc of CSCW 2004, ACM Press (2004), 182-191.

8. Dourish, P., and Bly, S. Portholes: Supporting Awareness in a Distributed Work Group. In Proc of CHI 1992, ACM Press (1992), 541-547.
9. Feldman-Barrett, L., and Barrett, D.J. An Introduction to Computerized Experience Sampling in Psychology. *Social Science Computer Review*, 19, 2 (2001), 175-185.
10. Fish, R.S., Kraut, R.E., and Chalfonte, B.L. The VideoWindow System in Informal Communication. In Proc of CSCW 1990, ACM Press (1990), 1-11.
11. Fish, R.S., Kraut, R.E., Root, R.W., and Rice, R.E. Evaluating Video as a Technology for Informal Communication. In Proc of CHI 1992, ACM Press (1992), 37-48.
12. Fogarty, J., Hudson, S., and Lai, J. Examining the Robustness of Sensor-Based Statistical Models of Human Interruptibility. In Proc of CHI 2004, ACM Press (2004), 207-214.
13. Fogarty, J., Hudson, S.E., Atkeson, C.G., Avrahami, D., Forlizzi, J., Kiesler, S., Lee, J.C., and Yang, J. Predicting Human Interruptibility with Sensors. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 12, 1 (2005), 119-146.
14. Fogarty, J., Ko, A.J., Aung, H.H., Golden, E., Tang, K.P., and Hudson, S.E. Examining Task Engagement in Sensor-Based Statistical Models of Human Interruptibility. In Proc of CHI 2005, ACM Press (2005), 331-340.
15. Fogarty, J., Lai, J., and Christensen, J. Presence versus Availability: The Design and Evaluation of a Context-Aware Communication Client. *International Journal of Human-Computer Studies (IJHCS)*, 61, 3 (2004), 299-317.
16. Gaver, W., Moran, T., MacLean, A., Lövstrand, L., Dourish, P., Carter, K., and Buxton, W. Realizing a Video Environment: EuroPARC's RAVE System. In Proc of CHI 1992, ACM Press (1992), 27-35.
17. Hatch, M.J. Physical Barriers, Task Characteristics, and Interaction Activity in Research and Development Firms. *Administrative Science Quarterly*, 32 (1987), 387-399.
18. Herbsleb, J.D., Atkins, D.L., Boyer, D.G., Handel, M., and Finholt, T.A. Introducing Instant Messaging and Chat in the Workplace. In Proc of CHI 2002, ACM Press (2002), 171-178.
19. Ho, J., and Intille, S.S. Using Context-Aware Computing to Reduce the Perceived Burden of Interruptions from Mobile Devices. In Proc of CHI 2005, ACM Press (2005), 909-918.
20. Horvitz, E., and Apacible, J. Learning and Reasoning about Interruption. In Proc of ICMI 2003, ACM Press (2003), 20-27.
21. Horvitz, E., Koch, P., and Apacible, J. BusyBody: Creating and Fielding Personalized Models of the Cost of Interruption. In Proc of CSCW 2004, ACM Press (2004), 507-510.
22. Hudson, J.M., Christensen, J., Kellogg, W.A., and Erickson, T. "I'd be overwhelmed, but it's just one more thing to do": Availability and Interruption in Research Management. In Proc of CHI 2002, ACM Press (2002), 97-104.
23. Hudson, S.E., Fogarty, J., Atkeson, C.G., Avrahami, D., Forlizzi, J., Kiesler, S., Lee, J.C., and Yang, J. Predicting Human Interruptibility with Sensors: A Wizard of Oz Feasibility Study. In Proc of CHI 2003, ACM Press (2003), 257-264.
24. Hudson, S.E., and Smith, I. Techniques for Addressing Fundamental Privacy and Disruption Tradeoffs in Awareness Support Systems. In Proc of CSCW 1996, ACM Press (1996), 248-257.
25. Iqbal, S.T., Adamczyk, P.D., Zheng, X.S., and Bailey, B.P. Towards an Index of Opportunity: Understanding Changes in Mental Workload During Task Execution. In Proc of CHI 2005, ACM Press (2005), 311-230.
26. Jones, E.E., and Nisbett, R.E. The Actor and the Observer: Divergent Perceptions of the Causes of Behavior. In Jones, E.E., Kanouse, H.H., Kelley, R.E., Nisbett, S.V., and Weiner, B. eds. *Attribution: Perceiving the Causes of Behavior*, General Learning Press, New York (1971), 79-94.
27. Kraut, R.E., Egidio, C., and Galegher, J. Patterns of Contact and Communication in Scientific Research Collaboration. In Proc of CSCW 1988, ACM Press (1988), 1-12.
28. Larson, R., and Csikszentmihalyi, M. The Experience Sampling Method. *New Directions for Methodology of Social and Behavioral Science*, 15 (1983), 41-56.
29. Mantei, M.M., Baecker, R.M., Sellen, A.J., Buxton, W.A., Milligan, T., and Wellman, B. Experiences in the use of a Media Space. In Proc of CHI 1991, ACM Press (1991), 203-208.
30. Milewski, A.E., and Smith, T.M. Providing Presence Cues to Telephone Users. In Proc of CSCW 2000, ACM Press (2000), 89-96.
31. Tang, J.C., Isaacs, E.A., and Rua, M. Supporting Distributed Groups with a Montage of Lightweight Interactions. In Proc of CSCW 1994, ACM Press (1994), 23-34.
32. Tang, J.C., Yankelovich, N., Begole, J., Van Kleek, M., Li, F., and Bhalodia, J. ConNexus to AwareNex: Extending Awareness to Mobile Users. In Proc of CHI 2001, ACM Press (2001), 221-228.