Information Extraction and eGovernment:
Crime Reporting and Investigative
Interviewing System

Completed Research Paper

Alicia Iriberri
California State University, Fresno
airiberri@csufresno.edu

Abstract
Crime statistics from the US Bureau of Justice and the FBI Uniform Crime Report show a gap between reported and unreported crime. For police to effectively prevent and solve crime, they require accurate and complete information about incidents. This article describes the evaluation of a crime reporting and interviewing system that witnesses can use to report crime incidents or suspicious activities anonymously while ensuring the information received is of such quality that police can use it to begin an investigation process. Findings of this evaluation show that the system is capable of producing accurate and complete reports by enhancing witnesses’ memory recall and that its efficacy approximate the efficacy of a human conducting a cognitive interview. The system is introduced as the first computer application of the cognitive interview and proposed as a viable alternative to face-to-face investigative interviews.

Introduction
Crime statistics from the US Bureau of Justice’s National Crime Victimization Survey (NCVS) and the FBI Uniform Crime Report Program (UCR) show a steady gap between reported and unreported crime from year to year. The NCVS, which collects information from victims of crime, showed that in 2012 only 44 percent of violent victimizations and 54 percent of serious violent victimizations, including domestic violence, were reported to police (Truman, Langton, Planty, 2013; FBI, 2012). These percentages do not change significantly from year to year and represent a considerable amount given the number of incidents that occur annually, 8.9 million violent and 19 million property crimes in 2012 alone. For police to effectively work on crime prevention and to solve crimes they require accurate and complete information about each incident. Research indicates that the information that witnesses provide to police is the primary determinant to solving crimes (Greenwood and Petersilia, 1975). However, many witnesses choose not to report, because of fear of repercussion, inconvenience, or distrust (Greenberg and Scott, 2004; Kidd and Chayet, 1984; Singer, 1988; Thomson and Langley, 2004).

We are developing a computer-based crime reporting and interviewing system that witnesses could use to report crime or suspicious activities to policy anonymously while ensuring the information is of such quality that police can begin an investigation. The crime reporting and interviewing system leverage natural language processing technology and cognitive interview techniques to present witnesses with a human conversational-style interface to ease the interviewing process and maximize memory recall. The system represents the first computer application of the cognitive interview. In this article we describe one empirical evaluation to test the efficacy of the computer system compared to the efficacy of a face-to-face (human) cognitive interview.

Theoretical Background

The Cognitive Interview
The cognitive interview, developed by Geiselman and Fisher (1985, 1997), is an interviewing technique that has shown to be effective in helping witnesses recall details of the events they witness and thus has
been recommended as the method of choice when conducting investigative interviews (Wells, Malpass, Lindsay, Fisher, Turtle and Fulero, 2000). This type of interview incorporates memory retrieval techniques that compared to traditional interviewing techniques (i.e., standard, questionnaire-like human investigative interviews) collect significantly more and more accurate information about incidents. In using these memory research-based techniques, police investigators induce witnesses to reinstate mentally the context of the event to be remembered, report in free and narrative form everything that they can remember, and state information starting with mental images that are richest in details and then continuing with less rich images until all details about the event to be remember are reported.

The efficacy of the cognitive interview has been tested extensively (Memon, Meissner, and Fraser, 2010). The variables used in the studies to compare the efficacy of interviewing methods are completeness (quantity of recall) and accuracy (quality or recall) (Geiselman and Fisher, 1997; Kohnken, Milne, Memon and Bull, 1999). Results of these studies report that the cognitive interview elicits an average of 41 percent (ranging from 15 to 147 percent) more correct details (completeness) compared to a standard police interview with 85 percent accuracy. The majority of these studies were conducted in laboratory settings with only a few conducted in the field (Colomb, Ginet, Wright, Demarchi and Sadler, 2013).

Although the cognitive interview is a highly effective technique, its use in the field is limited due to the limited training, time, and ability of police investigators (Hope, Gabbert and Fischer, 2006). Given that a cognitive interview may take police investigators at least 45 minutes and that right after a criminal incident occur investigators may need to interview various witnesses, oftentimes investigators choose to focus only on asking general, standard questions and on interviewing only key witnesses. This decision may result in the loss of important pieces of information (Dando, Wilcock, Milne, and Henry, 2008). Moreover when witnesses are not interviewed shortly after witnessing the incident, their memory of the incident may suffer from cross-contamination and decay (Hope, Gabbert, Fisher and Jamieson, 2014). Therefore, it becomes necessary to find avenues to leverage the benefits of using the cognitive interview techniques while alleviating police loads to ensure the flow of information between investigators and witnesses is not constricted and police can fully fulfill their duties.

**Information Extraction**

Information extraction is the application of natural language processing techniques to process written narratives, like texts, documents, articles or Web pages, to obtain fixed-format data about domain-specific entities included in the narratives. Once these data are extracted, they can be stored in database form and be used in queries for further analysis (Appelt and Israel, 1999).

Various techniques for processing narratives are used in information extraction. One of these techniques is rule-based, where example narratives are selected to identify such text portions as named entities, grammars, and text patterns and to create rules that model the grammar and text patterns identified in the example narratives. These rules are then incorporated to the processing of additional narratives to detect instances of the modeled portions of text and automatically extract and annotate these extractions. In the crime domain, named entities include people, locations, physical attributes, weapons, vehicles, criminal acts, and personal property. A text pattern in the crime domain may be for example, “brown eyes.” When this pattern is detected in a narrative, the information extraction process annotates it as eye:body part and brown:eyecolor, thus giving the processing system the ability to “understand” what is stated in the narrative.

Creating rules to extract information from narratives is a labor-intensive process, but it is the recommended method when the required level of precision is high and when example narratives available to creating and testing rules are limited. Rule-based information extraction systems often achieve high performance in extracting information (Appelt and Israel, 1999; Chiticariu, Li, and Reiss, 2013). Performance is usually measured in terms of recall, or number of matches found, and precision, accuracy of annotation. Common targets range from 50 - 90% recall and 70 - 90% precision depending on task complexity (Cardie,1997).

**The Reporting and Interviewing System**

The reporting and interviewing system used in this research emulates the tasks that a police investigator using the cognitive interview techniques would perform. First, the system asks witnesses to provide
general information about the crime, like date, time, and type of crime, and then to provide a written narrative of the incident they witnessed. Second, it extracts named entities and relevant facts from this narrative. Third, using the output of the initial extraction, the system assembles questions and designs an interviewing strategy, just like a cognitive interviewer does, and presents these questions to help witnesses recall facts that are missing in their report. Last, the system stores both complete narratives and annotated text in database form that can later be used to produce a written standard police report.

The system incorporates Web and database technology, Java-based proprietary code developed by the author and open-source code from the General Architecture for Text Engineering (GATE) program developed by researchers at the University of Sheffield (Cunningham et al., 2002). Using GATE’s information extraction tools and the crime domain rules specifically created to process witness narratives, in empirical tests the crime reporting and interviewing system shows that it is capable of extracting 96% of information from witness narratives with 100% precision (Iriberri and Leroy, 2007).

Evaluation of the Crime Reporting and Interviewing System

The purpose of the evaluation of the crime reporting and interviewing system is to measure the efficacy of the system in terms of its ability to collect accurate and complete crime reports from witnesses and to compare this efficacy to the efficacy of human interviewers conducting cognitive interviews. This evaluation would provide evidence to support the use of the system in the field to help alleviate police interviewing loads while ensuring the collection of complete and accurate reports. Iriberri and Navarrete, (2012) evaluated the system by comparing it to an example of existing Internet crime reporting alternatives that use an e-mail-like interface. In this evaluation the system outperformed the efficacy of the alternative method in terms of report completeness. This new evaluation would allow to benchmark the efficacy of the system by comparing it to a sound and more effective method.

In the evaluation of the system, participants were randomly assigned to one of two reporting methods: the crime reporting and interviewing system and a face-to-face (human) cognitive interview conducted by a trained cognitive interviewer. The efficacy of the system was compared to the alternative method expecting to find evidence to indicate that the system approximates closely the performance of human cognitive interviewers.

The user study was conducted in a laboratory setting. In this experiment, participants witnessed a staged crime depicted on a video clip and reported what they saw in the video. The following hypotheses represent the expected results:

Hypothesis 1 (H1):

Crime reports resulting from the use of the crime reporting and interviewing system are less complete than crime reports obtained by human cognitive interviews.

Hypothesis 2 (H2):

Crime reports resulting from the use of the crime reporting and interviewing system are less accurate than crime reports obtained by human cognitive interviews.

The study was a between subjects experimental design with one independent variable and two dependent variables. The output generated by the reporting method participants used was scored and means of each method were compared. The independent variable in the study design was the reporting method with two levels, the crime reporting and interviewing system and the human cognitive interview. The dependent variables were report completeness and accuracy. These variables are defined respectively as the number of correct items in the report (quantity of recall), and the number of correct items in the report over the sum of all items in report including correct items, incorrect items, and confabulated items (quality of recall).

To use as control variables, participants' demographic characteristics including age, gender, ethnicity, and educational level were collected using items on a questionnaire, which participants completed before the test. The questionnaire also included items relevant to the study materials.
Methodology

The user study and data collection was done on an individual basis in two separate offices. The offices were equipped with one computer with Internet connection, a widescreen monitor, and speakers. Individual participants were assigned to one condition. Participants assigned to use the system sat at a desk in front of the computer monitor throughout the study. Participants assigned to the human interview sat at a table where a computer monitor was located and one other chair was available for the cognitive interviewer to sit while conducting the interview and audio-recording the interview with a portable machine.

Participants watch a video clip depicting a staged crime. The video clip was two and a half minutes in length and is an excerpt of the commercial motion picture *As Good As It Gets* by James L. Brooks. This movie was rated by the Motion Pictures Association of America for audiences of 13 years of age or older (PG-13). In this video, participants observed a crime being committed by two suspects (Suspect 1 and Suspect 2) aided by an accomplice (Model) on a victim (Artist/Painter). The crime is a home robbery invasion and aggravated assault. The suspects take property from the painter's apartment while the model, who is posing for a portrait, distracts the painter. The painter is completing a full size portrait of the model. When the painter realizes (by his pet dog's actions) he is being robbed, he confronts the assailants and is assaulted by one of them, who hits him repeatedly with a coat rack. The two suspects and the accomplice flee the scene taking property items they had put in a sack. The portrait of the accomplice, which is a very accurate representation of the accomplice, is left behind at the scene.

After watching the video clip, participants using the system were asked to take a 20-minute brake and not to discuss with anyone the video they had watched. They were also asked to fill out the demographic questionnaire. After the brake, students using the system were introduced to the task they would have to perform and to the system interface using screenshots. Participants in the human interview were asked to grant permission to record the interview and were introduced to the trained cognitive interviewer, who began with a casual conversation.

Data Collection

In preparation for data collection, preliminary activities were performed, which included provisions to protect the study's internal validity. First, the stimulus video clip was coded and scored to identify relevant facts for a standard police report. One independent rater coded the stimulus. The video rater watched the video clip, extracted descriptive facts, and created a template grouping extracted facts in seven categories: `suspect-1, suspect-2, accomplice, weapon, property items, location, and criminal acts`. The total number of facts per category was tallied in the template. The template and the total number of facts represented the *gold standard* against which the resulting crime reports from each reporting method were compared and scored.

Second, three volunteered interviewers participated in a training session where they were introduced to the cognitive interview technique. The materials used in the training session were those created by Geiselman and Fischer and used in training police investigators various locations and certification programs.

Last, a pilot test of the study was run to ensure clarity and face validity of materials. This test served also as an opportunity to practice the study procedure and to verify the well-functioning of the reporting and interviewing system and for the human interviewers to fine-tune their skills. The pilot test was conducted a week prior to the study with eight volunteers. The crime reports created by these participants were not included in the study final dataset. Materials, procedures, and reporting methods were refined after the pilot test.

Procedure

Data collection activities were conducted in three different phases, presentation, testing, and data coding. In the presentation phase a facilitator (a) introduced participants to the study, (b) had participants watch the video in isolation on an individual bases, (c) surveyed the demographic profile of participants and their familiarity with the video clip actors in the movie; and (d) allowed participants to break for 20 minutes and asked them not to discuss the content of the video with anyone.
The 20-minute break was imposed to allow for a more naturalistic reporting situation. In real crime reporting situations, time passes by before a witness reports incidents to police. During the delay, participants were set to converse with a study confederate to ensure that they would not have the opportunity to replay the video in their minds trying to remember every detail or to discuss its content with anyone.

In the test phase the facilitator (a) assigned participants randomly to each reporting method; (b) instructed participants on the use of the reporting methods using screenshot slides; (c) instructed participants to begin their reports using the assigned condition; and (d) debriefed participants when they completed their reports and asked them not to discussed the study with anyone likely to participate in the study.

During the data coding phase the facilitator (a) printed all reports stored in the crime reporting and interviewing systems, (b) transcribed all recorded interviews, (c) submitted resulting reports to be coded and scored, (c) verified the accuracy and completeness scores of each report, and (d) coded demographic questionnaire results.

All reports and questionnaires were identified with a participant’s numeric code, only the primary investigator knew the names of participants and the reporting method they used. An independent rater coded and scored the resulting crime reports. The rater scored reports by hand using color-coding (one color per crime video’s category), and tallied totals using the gold standard template created by the video clip coder.

Once reports were coded and scored, two types of statistics were calculated to allow for comparison between the two experimental conditions. These statistics were descriptive statistics and chi-squares to obtain a profile of participants’ demographic characteristics and verify random assignment, and one-way ANOVAs to test the study’s hypotheses for the two dependent variables completeness and accuracy.

**Results**

Resulting reports were prescreened to ensure content validity yielding 46 usable reports out of the 70 reports received. In this process of screening reports, reports from witnesses who stated in the pre-test questionnaire that they were familiar with the movie and/or the actors were dropped. Of the 46 usable reports, 23 resulted from the crime reporting and interviewing system and 23 from human interviews. Three interviewers conducted the 23 interviews. One produced 10 reports, the second produced six reports, and the third produced seven reports.

**Participants**

Participants in the study were sophomore students enrolled in various sections of the Management Information Systems course in the College of Business Administration of State University. These students majored in accounting; finance, real state and law; international business; e-business; marketing management; computer information systems; management and human resources; and technology and operations management, and all were computer literate. College students in their second year were selected to allow for a controlled sample in terms of educational level and verbal ability. The aim was to have a representative sample of the US adult population, where 52% have achieved some college education (2000 US Census). We consider this population to be the target users of Internet crime reporting methods.

The demographic profile for the 46 participants who provided usable reports showed that the sample was representative of the target population. Seventy-three percent of the participants were between 18 and 25 years old, 51 percent were males, and 86 percent had attained some college education. Statistics also indicated that the sample was ethnically diverse with all major ethnic background (the resulting ethnicity distribution is representative of the student population of the study site). Chi-squares for demographic variables were calculated to verify that participants were equally distributed among reporting conditions.
**Coding and Scoring of Reports**

The 46 witness reports each with an average length of four pages were coded. Scores for each report were calculated by tallying the presence of descriptive items that relate to the video clip gold standard. Three totals were obtained, number of correct items (those that match items in the gold standard), number of incorrect items, and number of confabulated items (those not present in the gold standard).

**Completeness**

One-way ANOVA (one-tailed) was used to test for completeness (number of correct items in report) among the two reporting methods. Using an alpha level of 0.05, significant differences in mean completeness were observed. Table 1 shows completeness data and ANOVA main effect. Completeness differed significantly across the two methods $F(1, 45) = 7.019, \eta^2 = 0.155, p = .011$.

The results show that the crime reporting and interviewing system elicited 51 correct descriptive items from witnesses’ memory and the human interview elicited 70 correct items. This indicates that the crime reporting and interviewing system elicits 26 percent less than the human interviewer. Thus, results for completeness suggest that the human interview outperformed the crime reporting and interviewing system in completeness. The direction of the results of the comparison with human interview was as expected.

<table>
<thead>
<tr>
<th>Reporting Method</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Human</strong></td>
<td>23</td>
<td>69.1</td>
<td>21.9</td>
<td>4.6</td>
</tr>
<tr>
<td><strong>System</strong></td>
<td>23</td>
<td>51.4</td>
<td>23.3</td>
<td>4.9</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sum of Squares</th>
<th>Df</th>
<th>Mean Square</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between Groups</td>
<td>1</td>
<td>3583.391</td>
<td>7.019*</td>
<td>.011</td>
</tr>
<tr>
<td>Within Groups</td>
<td>44</td>
<td>510.534</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>45</td>
<td>26046.870</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Effect is significant at the 0.05 level

Table 1. Completeness Data by Reporting Method and ANOVA Main Effect

<table>
<thead>
<tr>
<th>Reporting Method</th>
<th>Mean Correct</th>
<th>Mean Incorrect</th>
<th>Mean Total in Report</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Human</strong></td>
<td>69.1</td>
<td>2.5</td>
<td>71.6</td>
</tr>
<tr>
<td><strong>System</strong></td>
<td>51.4</td>
<td>3.3</td>
<td>54.7</td>
</tr>
</tbody>
</table>

Table 2. Mean of Items in Report by Reporting Method

**Accuracy**

One-way ANOVA (one-tailed) was used to test for accuracy (number of correct items over the sum of correct items plus incorrect items plus confabulated items) between the two reporting methods. Some incorrect (incorrectly remembered) and confabulated items (not present in the video) were obtained from participants across the two methods. Incorrect and confabulated items were added together because they represented less than five percent of total items in each report (Table 2). The crime reporting and interviewing system (M=3.3) produced more incorrect items than the human interview (M=2.5).

Using an alpha level of 0.05, significant differences in mean accuracy of reports were observed. Table 3 shows accuracy data and ANOVA main effect. Accuracy also differed across the two methods $F(1, 45) = 8.053, \eta^2 = 0.155, p = .007$. 
The results show that the human interview elicited items from witness memory with 97 percent accuracy and the crime reporting and interview system elicited items with 94 percent accuracy. This suggest that 97 percent of the information the human interview elicits is correct and that 94 percent of the information the system elicits is correct.

<table>
<thead>
<tr>
<th>Reporting Method</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human</td>
<td>23</td>
<td>.968</td>
<td>.023</td>
<td>.005</td>
</tr>
<tr>
<td>System</td>
<td>23</td>
<td>.938</td>
<td>.044</td>
<td>.009</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between Groups</td>
<td>.010</td>
<td>.010</td>
<td>8.053*</td>
<td>.007</td>
</tr>
<tr>
<td>Within Groups</td>
<td>.055</td>
<td>.001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>.065</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Effect is significant at the 0.05 level

Table 3. Accuracy Data by Reporting Method and ANOVA Main Effect

Discussion

The results obtained in this user study support the hypotheses. Differences among the methods in report completeness were significant. The human interviewer outperformed the crime reporting and interviewing system in report completeness, with 69 correct items vs. 51 correct items. That is, the face-to-face (human) cognitive interview elicited 26 percent more correct item than its computer application.

Evaluations of the cognitive interview report that, in term of completeness, the face-to-face cognitive interview outperforms other face-to-face interview methods, the standard police interview and the structured interview, by an average of 41 percent (Kohnken et al., 1999; Memon et al., 2010). Similarly, the computer crime reporting and interviewing systems outperformed a representative example of existing computer-based reporting systems by 143 percent (Iriberri and Navarrete, 2012).

In terms of accuracy, both the face-to-face method and the computer crime reporting and interview system achieved high accuracy, with 97 percent for the cognitive interview vs. 94 percent for the system. Koehnken et al.’s meta-analysis of 55 evaluations of the cognitive interview indicates that accuracy is affected by age of participants, delay between viewing stimuli and reporting the incident, and differences in the laboratory where the studies were conducted. In this study, the short delay is likely the reason for the methods to achieve high accuracy levels. High levels of accuracy are not rare since the cognitive interview is a sound technique. Other studies of the cognitive interview have reported anywhere from 76 percent to 100 percent accuracy.

The results of this study thus show that the system approximates the efficacy of the human cognitive interview and is a potential substitute when face-to-face interviews are not possible or feasible. In turn, when used the system could alleviate both police interviewing loads and witness’ reporting concerns.

Conclusions

Findings of the evaluation of the crime reporting and interviewing system show that it can collect accurate and complete information and that its performance approximates the performance of a human investigator conducting an interview of a witness. These results lead to two conclusions on the efficacy of the crime reporting and interviewing system. First, the computer crime reporting and interviewing system enhances witness communication and memory retrieval in a way similar to the human cognitive interview. In terms of completeness, the human interviewer outperformed the crime reporting and interviewing system. In terms of accuracy, the human interview also outperformed the computer crime reporting and investigative system. However, the direction of these results was as expected given that the comparison is between a human and a computer. However the relevance of this finding is the fact that the system does approximate the performance of the human interview and may serve as an appropriate
substitute when face-to-face cognitive interview are not feasible. Iriberri and Navarrete (2010, 2013) evaluated the system comparing it to existing Internet reporting alternatives and the system outperformed these alternatives. In perspective, the system is an improvement over existing alternatives and with further refinements we expect it may approximate the performance of the human interview even closer.

Second, natural language processing technology allowed the crime reporting and interviewing system to apply successfully the cognitive interview memory-enhancing guidelines. Using these investigative interviewing techniques, the crime reporting and interviewing system is capable of eliciting accurate and complete witness testimonies. Computer interviews using natural language interfaces are promising technologies to offer interactive and usable interfaces in e-government services to the public and may warrant further research.

Limitations

The evaluation of the crime reporting and interviewing system was conducted in a laboratory setting, which limits the generalizability of result to the real world. Colomb et al. (2013) report on the limited number of evaluations of the cognitive interview conducted in the field and the challenges that field settings bring to cognitive interview research, but concludes that the benefits of the cognitive interview seems to generalize in a real world context.

In this evaluation of the crime reporting and interviewing system, external validity threats were anticipated and reduced when possible. These threats include (a) possible ceiling effect for reporting methods’ accuracy; (b) un-naturalistic reporting situation due to participants acting as witnesses of a crime depicted in a video clip. The alternative methods achieved accuracy levels in the range of 94 to 97 percent. Although these results are common in studies on the evaluation of the cognitive interview when compared with other alternative interviewing methods (i.e., standard interview), the results for the system’s accuracy should be taken with caution. Longer delays between the stimulus and reporting the crime incident may yield more realistic results. The circumstances in which participants witnessed a crime and filed a report in this study are by no means similar to real witnesses reporting on real crimes. To advance generalizability of the results of this study, field studies should be conducted. This study however allowed for a comparison of the efficacy of the system compared to the efficacy of a face-to-face (human) interview under similar circumstances.

Implications

Crime statistics indicate that millions of crimes are committed every year, but less than half are reported to police (Truman et al., 2013). Unwillingness to report, driven by inconvenience or fear, is cited as the primary reason for unreported crime. Singer (1988) advocated the creation of alternatives to mediate between police’s goals and victims’ fears. Accordingly, law enforcement agencies launched Internet crime reporting as an alternative to human reporting. With an offer for convenience and privacy, the crime reporting and interviewing system is a promising example of Internet crime reporting that may address unreported crime while enduring complete and accurate crime reports. This represents a clear advantage over existing Internet reporting implementations and is an indication of the value of using human conversational-style interfaces to elicit more information from citizens. The contribution of this study is the first computer application of the cognitive interview. Further studies will test the performance of the crime reporting and interviewing system in the field with more naturalistic settings.

REFERENCES


