

Psychologically based Virtual-Suspect for Interrogative Interview Training*

Moshe Bitan

Department of Computer Science
Bar-Ilan University, Israel
moshe.bitan78@gmail.com

Galit Nahari

Department of Criminology
Bar-Ilan University, Israel
naharigalit@gmail.com

Zvi Nisin

zvinisin@gmail.com

Ariel Roth

Department of Computer Science
Bar-Ilan University, Israel
roth.ariel.phil@gmail.com

Sarit Kraus

Department of Computer Science
Bar-Ilan University, Israel
sarit@cs.biu.ac.il

Abstract

In this paper, we present a *Virtual-Suspect* system which can be used to train inexperienced law enforcement personnel in interrogation strategies. The system supports different scenario configurations based on historical data. The responses presented by the Virtual-Suspect are selected based on the psychological state of the suspect, which can be configured as well. Furthermore, each interrogator's statement affects the Virtual-Suspect's current psychological state, which may lead the interrogation in different directions. In addition, the model takes into account the context in which the statements are made. Experiments with 24 subjects demonstrate that the Virtual-Suspect's behavior is similar to that of a human who plays the role of the suspect.

Introduction

One of the most important tasks of a police officer is to interrogate suspects. The main role of the police interrogation is to encourage the suspect to inadvertently incriminate himself and to collect further information regarding the case. Often, the physical evidence available in a given investigation is scarce, which makes the interrogation stage much more critical. Therefore, police cadets undergo rigorous interrogation training.

One of the leading training techniques is one-on-one interrogation simulation sessions. In these personal sessions, a trained instructor conducts an interrogation simulation, with the trainee acting as the investigator. In order to prepare the inexperienced law enforcement personnel for real-world investigations, the instructor must develop a scenario based on real cases. Furthermore, the instructor or a hired actor plays the role of the suspect and portrays different personalities based on the corresponding scenario. The interrogation uses the *Investigative Interviewing* method and is based on an interrogation plan which begins with a relaxed atmosphere of acquaintance, then the suspects version of the event is obtained and his/her alibi is verified, and finally tough questions are introduced and in some cases accusations are made

*This work was supported by the LAW-TRAIN project that has received funding from the European Union Horizon 2020 research and innovation program under grant agreement No 653587 and the ERC (grant No. 267523).
Copyright © 2017, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

against the suspect. Interestingly, interrogators often cycle back to one or more of these steps according to the suspect's state of mind. This training technique has long served its goals and has proven efficient and effective. Unfortunately, this time consuming technique requires experienced instructors or actors. Furthermore, the training sessions are carried out on a one-trainee-at-a-time basis, and are therefore expensive.

Virtual-Suspect systems may offer a solution for this. The advantages of a computer interrogation training system are fairly straightforward. First, the system can provide a large number of scenarios and satisfy greater control over the Virtual-Suspect's personality. Second, multiple cadets can train simultaneously at their convenience with an instructor monitoring their progress. Lastly, detailed reports can be provided as well as transcripts of an interrogation which can facilitate instructors in monitoring the training.

In this paper, we present a Virtual-Suspect system. The system was developed in collaboration with criminology researchers, experienced criminal psychologists and the National Police Department. The system supports different scenario configurations based on real cases. The responses presented by the Virtual-Suspect are selected based on the psychological state of the suspect, which can be configured as well. Furthermore, each interrogator's statement affects the Virtual-Suspect's psychological state, which may lead the interrogation in different directions. An experiment was conducted comparing the system's responding mechanism with that of a human instructor. The experiment was divided into two phases. First, interrogation simulations were conducted either using a human instructor, our system or a baseline random system. Second, participants were asked to read the transcripts of these simulations and answer a series of questions. The results suggest that humans have difficulty differentiating between simulations generated by our system and those of a human instructor. To conclude, the cost efficient *Virtual-Suspect* system can be used around the clock to train inexperienced law-enforcement personnel in a variety of realistic interrogations.

Related Work

The effectiveness of Virtual-Humans in training systems has been studied extensively in computer science (Hubal and Frank 2001; Kenny et al. 2007; Parsons et al. 2008; ?). An-

derson et al. (Anderson et al. 2013) presented the *TARDIS*, an interview coaching system for training scenario-based interactions. The framework incorporates a *Non-Verbal Behavior Analyzer (NovA)* sub-module which can recognize several lower level social cues, e.g. hands-to-face, postures, leaning forward or backward, among others. Luciew et al. (Luciew, Mulkern, and Punako 2011) presented an immersive interrogation learning simulation specifically designed for training law enforcement personnel in interviewing children who were victims of sexual abuse and interrogating suspects on that matter. The study puts a strong emphasis on the interpretation of nonverbal behavior by police officers during an interview or interrogation. Lin et al. (Lin, Oshrat, and Kraus 2009; Lin et al. 2014) studied whether agents can be used for training people in negotiation and interviewing in role playing simulations. The study showed that automated agents that emulate human behavior in such settings can improve people's performance at least as well as when the other role is played by a human. The Virtual-Suspect system presented in this paper builds on these findings and provides a Virtual-Agent that plays the role of a suspect in the same manner as a person would.

Other researchers have focused their attention on the Virtual-Suspect's psychological model. Roque and Traum (Roque and Traum 2007) categorized three compliancy levels: compliant, reticent and adversarial. A compliant suspect provides semi-useful information when asked. A reticent suspect provides neutral information and evades any questions about sensitive topics. An adversarial suspect provides deceptive or untruthful responses. Similar to this compliancy categorization, the Virtual-Suspect's internal-state as presented in this paper integrates a continuous compliancy parameter which, in combination with other parameters, determines the suspect's response. Olsen et al. (Olsen 1997) presented a police interrogation simulation system teaching police cadets to build a rapport with the suspect while maintaining professionalism, to listen to verbal cues and to detect important changes in both verbal and nonverbal behavior. The Virtual-Suspect, a male loan officer, is accused of stealing from an ATM. The presented response model is based on the suspect's internal-state, comprising of the suspect's mood and the rapport that has been established between the suspect and interrogator. The system presented in this paper extends Olsen's internal state model and introduces a three-dimensional internal state. Furthermore, the system supports multiple configurable interrogation scenarios typically based on real cases. Lastly, Bruijnes et al. (Bruijnes et al. 2014) presented a Virtual-Suspect response model that can portray a variety of unique personalities. The system is based on static personality parameters as well as a dynamic interaction state. Similar to the experiment presented in this paper, Bruijnes conducted an experiment measuring a human being's ability to distinguish between different Virtual-Suspect personalities.

Personality traits have been used in automated agents that negotiate with people (Lin and Kraus 2010). Such traits were implemented, for example, in *Diplomat* (Kraus and Lehmann 1995), where the agent played the game *Diplomacy* with people. It was possible to change the traits at

the beginning of each game, which affected the behavior of *Diplomat* between games. It allowed *Diplomat* to change its personality from one game to another and to act non-deterministically, which contributed to its success in the game. However, the traits were not changed during the game as they were with our VS. On the other hand, *Diplomat* had a limited learning capability which allowed it to try to estimate the personality traits of its rivals. The VS does not model the interrogator directly but rather the interrogators actions influence the VSs internal state during the interview.

Methodology

The *Virtual-Suspect* (VS) simulation system proposed in this paper enables the simultaneous simulation of multiple interrogation cases and supports different VS personality configurations. Furthermore, a simulated interrogation can be unfolded in many different directions depending on the trainee's behavior. In order to facilitate this flexibility, the system is comprised of two major components: The first component is the *Simulation Configuration* and the second component is the *Virtual-Suspect Response Model*.

Simulation Configuration

The *simulation configuration* provides a high level of control over the simulated interrogation. First, the interrogation scenario configuration enables the development of complex investigation simulations, typically based on real cases. Second, the system supports a customizable template database of statements and responses. Third, the Virtual-Suspect's personality, which is based on Eysenck's PEN Model (Eysenck 1990), can be configured. These three levels of control enable the interrogation trainer to simulate a wide range of interrogation scenarios.

1. Interrogation Scenario Configuration

The scenario configuration is implemented via a personal information database and an event database. The personal information database stores the VS's *age, marital status, spouse, children, last known address, occupation, income, place of employment, known acquaintances* and any other relevant personal data. The event database contains information about occurrences that may or may not have happened to the VS. These occurrences contain multiple types of information which relate to the occurrence, that is the *location, time, date, activity, participants, objects* and *means of transportation*. For example, on the 24th of December 2013 at 8:30 pm, the VS and his wife dined at their residence.

An event can be either truthful or false. A truthful event is an event that actually happened to the VS. On the other hand, a false event is an occurrence that the VS may use in order to give a deceptive response, and it is usually associated with providing an *Alibi* or *Legal Access*. Therefore, events are labeled with one of the following mutually exclusive labels: *Criminal, Alibi, Legal Access* or *Neutral*. A *Criminal* event is the offense details allegedly committed by the VS. An *alibi* event is a defense used by the VS attempting to prove that he or she was in some other place at the time that the alleged

Example #1: The Burglary Scenario

Investigator: What do you do at work?
Suspect: I am a welder.
Investigator: What hours do you work?
Suspect: We finish we leave.
Investigator: Where did you work?
Suspect: I worked at A. A. Metals as a welder.
Investigator: What did you do on the evening of Jan 1, 2013?
Suspect: Don't remember
Investigator: Why are you being evasive and not answering my question?
Suspect: It's not that I am not answering, I truly don't remember.
Investigator: The earrings found in your house are identical to the ones reported stolen. How could that be?
Suspect: We probably bought them in the same store.
Investigator: Did you pay for the earrings in cash or by credit?
Suspect: My head hurts.

Example #2: The Drug Trafficking Scenario

Investigator: Do you know Chris?
Suspect: Yes.
Investigator: Could you elaborate?
Suspect: I know him from the Gym, we work out.
Investigator: What did you do the morning of October 10, 2014?
Suspect: I went for a ride on my bicycle.
Investigator: Could you elaborate about that?
Suspect: I just took my bike and rode in the fields.
Investigator: Who was with you on that day who can verify this?
Suspect: What's the difference?
Investigator: I am telling you that witnesses saw you at the airport on the morning of October 10, 2014, were you there?
Suspect: I do not know, I really do not know what I did then.
Investigator: In the search of your apartment receipts for a flight were found. Explain how you have them.
Suspect: What do you mean? It is mine. Don't you see that it is in my house?

Figure 1: Simulation Transcripts

offense was committed. A *Legal Access* event is another defense used by the VS attempting to provide a legal explanation for its presence in the location of the alleged offense, association with an alleged accomplice to the offense or access to alleged stolen items. Lastly, a *Neutral* event is any event other than the above mentioned events. This event database models the long-term memory of the VS. However, in real-world cases, the police investigators are provided with only a partial view of what actually happened. To simulate this, at the onset of a simulation, the trainee is provided with a case file which consists of the information known to the police as well as the collected forensic evidence.

In real-world investigations, some topics can trigger a more emotional response than others. In order to simulate this behavior, a dedicated label was introduced to the system. The *hot* label is an indication that a personal detail, an event or any detail pertaining to an event has a profound effect on the VS when asked about the matter. It is important to note that an event or an event detail can be labeled as *hot* regardless of the event label. For example, if the VS's spouse's information is labeled as *hot*, an inquiry about a *Neutral* event where the VS and his wife dined at their residence might have a profound effect on the VS's state of mind. A more in-depth explanation of the *hot* label and its effects is provided later in the "*Response Model*" section.

To demonstrate the flexibility of the Interrogation Scenario Configuration, two scenario configuration examples are presented. The first example is the scenario used in the experiment. The second example was configured for the purpose of demonstrating the ease of use of the system. Furthermore, the complex scenario was configured in less than two hours. Simulation transcripts, conducted with these two scenario configurations, can be viewed in Figure 1.

Example 1 The interrogation scenario chosen for the experiment is based on an actual burglary case from early 2013. A 46-year-old married welder was charged with breaking and entering into a private residence and stealing a pair of valuable earrings and a laptop computer. Forensic investigation led to the finding of the suspect's fingerprints on a window ledge. As a result, a search warrant was granted, leading to the discovery of the earrings hidden in the suspect's residence. Finally, the suspect was arrested and

escorted to the interrogation room.

Example 2 The second interrogation scenario is also based on a real case from late 2014. Bob, a 28-year-old married delivery man was charged with drug trafficking. The suspect's contact information was found on a sim-card belonging to Chris, a second offender. Chris was arrested at the airport as he was trying to leave the country. An on-site search led to the discovery of 2Kg of an illegal substance in his possession. Furthermore, witnesses have testified to seeing Bob's car at the airport when he allegedly purchased the plane tickets for Chris. A further forensic investigation led to the finding of the suspect's fingerprints on a ticket purchasing booth. As a result, a search warrant was granted, leading to the discovery of the ticket receipts hidden in the suspect's residence. Finally, the suspect was arrested and escorted to the interrogation room.

2. Interrogative Interviewing The interrogation simulation is based on a standard Interrogative Interviewing model. Unlike other systems that support a single static scenario (Olsen 1997; Bruijnes et al. 2014), the VS system supports multiple configurable scenarios. Therefore, the VS system provides the ability to configure statement and response templates. Each statement or response template is comprised of a static text portion and a dynamic input fields portion. The input fields portion is filled out by the investigation trainee during the simulation. For example, "Where were you on 01/01/2013?" is an instance of the "Where were you on [Date]?" statement template. During the simulation, the trainee selected the statement template from the templates database and entered the date '01/01/2013'.

The association of a statement template to a response template is a *Many-to-Many* relation. In other words, a statement can have multiple response templates associated with it. Symmetrically, a response template can be associated with multiple statement templates. For instance, the statement "How are you?" can be associated with "I am feeling well" as well as with "I am a little agitated". Likewise, the response template "I don't remember" can be associated with the statement template "Where were you on [Date]?" as well as to "Where did you purchase these [Objects]?".

During a simulation, the trainee selects a statement from

the statement-response templates database, fills in the input fields and sends the statement to the VS. Consequently, the VS's response model extracts all associated responses from the templates database, fills in the required input fields from the databases, and selects an appropriate response from the associated responses set.

As described above, the *hot* label indicates that an information detail (i.e. a field) has a profound effect on the VS when asked about the matter. To simulate this behavior, the *hot* label is propagated at runtime to the Interrogative Interviewing model. A response or a statement is marked as *hot* if one or more of its input field values is labeled *hot* in the Scenario Database. In addition, a statement is also marked *hot* if any of its associated responses is marked as *hot*. Lastly, if a statement is not marked as *hot* it is considered to be *cold*. Formally, the indicator function δ^{hot} indicates that a given statement q is marked as *hot*:

$$\delta^{hot}(q) = \begin{cases} 1, & \exists(hotfield) \in \cup_{r \in R_q} F^r \cup F^q \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

Where q denotes the statement, r denotes a response, and R_q is the associated responses set, F^q and F^r are the set of input fields for the templates q and r , respectively.

A statement directly affects the internal-state of the VS. However, not all statement's subjects have an equal effect on the psychological state of the suspect. Therefore, the weight of the effect on the internal-state of each statement template can be tuned individually during the configuration. For each statement template q , two weight vectors are defined. The first, denoted by w_q^{hot} , determines the effect when the statement is marked as *hot*. The second, denoted by w_q^{cold} , determines the effect when the statement is marked as *cold* (i.e. not *hot*). An effect weight component can be either 1, -1, or 0 representing a positive, a negative and a neutral effect, respectively. Since the internal-state is represented using a three-dimensional vector, these weight vectors are three-dimensional as well. Formally, the vectors can be described as: $w_q^{hot}, w_q^{cold} \in \{-1, 0, 1\}^3$. These vectors were determined separately by two experts. In case of a disagreement, a discussion was made to resolve it.

3. Personality Profile The personality profile consists of two intertwined components. The first is the VS's *Internal-State* which is based on Eysenck's *PEN* Model of personality (Eysenck 1990). The second is the volatility of the VS's personality, that is, the degree to which a statement affects the internal-state.

The VS's internal-state is based on Eysenck's *PEN* Model of personality (Eysenck 1990). More specifically, Eysenck's *PEN* Model consists of the *Psychoticism*, *Extraversion* and *Neuroticism* personality traits where *Psychoticism* refers to a personality pattern typified by aggressiveness and interpersonal hostility. *Extraversion* tends to be manifested by outgoing, talkative, energetic behavior, whereas introversion is manifested by a more reserved and solitary behavior. *Neuroticism* is a fundamental personality trait in the study of psychology characterized by anxiety, fear, moodiness, worry, envy, frustration, jealousy, and loneliness. Individuals who score high for neuroticism are more likely than the

average human to experience such feelings as anxiety, anger, envy, guilt, and depression. Formally, the internal-state is a three-dimensional vector denoted by $s = (s_1, s_2, s_3) \in \mathbb{R}^3$ where the s_1, s_2 and s_3 components correspond to the *Psychoticism*, *Extraversion* and *Neuroticism* personality traits, respectively. The initial value of the internal-state vector, denoted by s_0 , is configured by the interrogation simulation supervisor (i.e. trainer) and is an integral part of configuring an interrogation simulation.

The internal-state changes during an interrogation simulation in order to reflect the effect that the interrogator's statements have on the VS. The degree to which the internal-state varies is determined by the second component, that is, the personality *Volatility* parameter. More specifically, the *Volatility* parameter, denoted by σ , is a non-negative, three-dimensional weight vector $\sigma \in \mathbb{R}_+^3$ that determines the pace at which the internal-state changes for any given statement. For instance, configuring the *Volatility* parameter, σ , to a value of (1, 0, 0) will cause the interrogator's statements to only effect the *Psychoticism* component of the internal-state vector. On the other hand, setting the σ parameter to a high value such as (3, 3, 3) will cause the internal-state vector to change dramatically after every statement. Consequently, the VS will react in an erratic and unstable manner. Although this *Volatility* parameter is configured in the Simulation Configuration part, its main function is in updating the internal-state vector which is part of the Response Model (see the "*Internal State Vector Update Mechanism*" section).

Personality Profile Presets In order to accelerate the simulation configuration, the *Personality Profile Preset Module* was integrated into the system. The interrogation simulation supervisor can easily choose a predefined personality profile from a wide range of presets. In addition, the supervisor can define custom presets and save them for later use. The preset consists of the initial value of the internal-state vector s_0 and the *Volatility* parameter σ . For example, the preset "*Moderately Calm Suspect*" is configured to $s_0 = (0, 0, -3)$ and $\sigma = (0.5, 0.5, 0.5)$. Another preset example is the "*Unstable Suspect*", which is configured to $s_0 = (-3, 3, 3)$ and $\sigma = (1, 1, 1)$. The predefined presets were configured by two experts and were tested over multiple simulations.

Response Model

The Virtual-Suspect *Response Model* generates the VS's response for a given inquiry, i.e. an investigator's statement. The model is comprised of four cognitive components. The first two components are the long- and short-term memory components. The long-term memory gives the VS access to the Interrogation Scenario database to recall personal information and past events. Equally important, the short-term memory gives the VS the ability to respond to follow-up statements. The third component, and arguably the most important, is the VS's internal-state update mechanism which determines the way an investigator's statement affects the VS's internal-state. Finally, all of the above components are combined to produce the response of the VS.

As can be viewed in Figure 2, once a statement q is selected by the trainee, the response model generates a re-

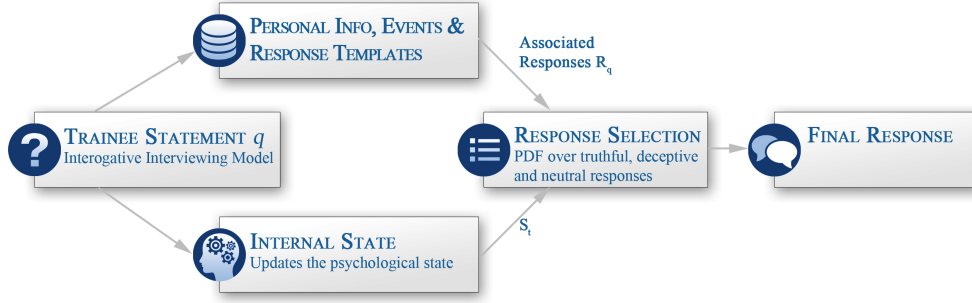


Figure 2: Response Model

sponse using the following steps. First, the model extracts the associated responses set R_q from the *Interrogative Interviewing Database*. Second, the long- and short-term memory are used to populate all input fields for each response in the associated responses set R_q . Next, the *Response Model* updates the internal-state vector. Finally, the updated internal-state vector is used to select the most appropriate response from the associated responses set R_q .

1. Long- and Short-Term Memory The main purpose of the long- and short- term memory is to populate the fields for the associated responses set R_q . Given a statement q , the *Response Model* determines if the statement is a follow-up statement, refers to a new event or is a generic statement (e.g. how are you feeling today). If the statement refers to an event, the model extracts all relevant information from its long-term memory (i.e. the Scenario Database) and stores it in its short-term memory. Lastly, the *Response Model* populates all input fields for each response in the associated responses set R_q using the event information stored in the short-term memory and the personal information database.

2. Internal State Vector Update Mechanism The *Internal-State Vector Update Mechanism* is executed for every statement. However, not all statement's subjects have an equal effect on the psychological state of the suspect, as stated above. During a simulation, for every statement q at time t the internal-state vector is calculated using the following equation:

$$s_t = s_{(t-1)} + \sigma \cdot (\delta^{hot}(q) \cdot w_q^{hot} + (1 - \delta^{hot}(q)) \cdot w_q^{cold}) \quad (2)$$

Where σ is the *Volatility* parameter, $\delta^{hot}(q)$ indicates if a statement q is marked as *hot* and w_q^{hot} , w_q^{cold} are the statement effect weight vectors. It is important to note that the *Volatility* parameter σ is configured in the Personality Profile configuration phase and remains constant for the duration of the simulation.

3. Response Selection The final phase in the response model is the response selection phase. In this phase the internal-state vector is used to select the output response, denoted by r^* , from the associated responses set R_q . More specifically, the internal-state determines what type of response to select: a truthful, deceptive or neutral response.

As can be seen in Figure 3, the response selection process consists of four steps.

First, the associated responses set, R_q , is divided into three subsets: : separate subsets with truthful, false and neutral responses. The truthful responses subset, denoted by R_q^t , contains all responses that relate to truthful events, that is events that actually happened to the Virtual-Suspect. The false responses subset, denoted by R_q^f , contains all deceptive responses, that is responses related to *Alibi* or *Legal-Access* events. The neutral responses subset, denoted by R_q^n , contains all general responses, for instance, "I do not recall", "I cannot remember" and "Let me be".

In the second step, the internal-state vector components are color-coded and then discretized to the *Mental Integrity* vector, denoted by s' . First, the value range of each internal-state vector component is divided into three sections, denoted by I_g , I_o and I_r , where each segment is color-coded to *green*, *orange* and *red*, respectively. The color-coding expresses the level of the Virtual-Suspect's *Mental Integrity*. The *green* color-coded section, I_g , indicates that the Virtual-Suspect is mentally stable. The *orange* color-coded section, I_o , indicates that the Virtual-Suspect is moderately stable. Lastly, the *red* color-coded section, I_r , indicates that the Virtual-Suspect's mental integrity is compromised. Formally, let s be the internal-state vector, and the *Mental Integrity* vector $s' \in \{0, \dots, 3\}^3$ components can be calculated as follows:

$$\begin{aligned} s'_1 &= \sum_{i=1}^3 \delta(s^i \in I_g), & s'_2 &= \sum_{i=1}^3 \delta(s^i \in I_o) \\ s'_3 &= \sum_{i=1}^3 \delta(s^i \in I_r) \end{aligned} \quad (3)$$

where δ is the standard indicator function (i.e. equals 1 if the condition is true, otherwise equals 0). The sections were carefully calibrated by police psychologists. For *Psychoticism* and *Extraversion*, the sections were set to $I_g = [-3, 3]$, $I_o = (-\infty, 3) \cup (3, 5]$ and $I_r = (5, \infty)$. For *Neuroticism*, the sections were set to $I_g = [-3, 3]$, $I_o = [-5, -3) \cup (3, \infty]$ and $I_r = (-\infty, -5)$. Figure 3 provides a visual representation of the segments.

The main objective of the Virtual-Suspect is to lead the interrogator to believe he or she has nothing to do with

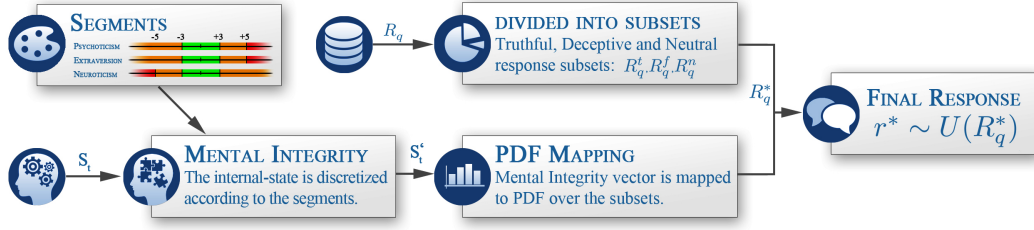


Figure 3: Response Selection

the alleged offense. Therefore, if the Virtual-Suspect is guilty it will try to deceive the interrogator when asked about topics related to the *Criminal* event. However, if the Virtual-Suspect’s mental integrity is compromised, there is a higher probability that the Virtual-Suspect will inadvertently choose a response that does not serve its goals. Furthermore, when asked about a *hot* topic that is not related to the *Criminal* event, the Virtual-Suspect will try to avoid the question and there is a high probability that it will respond with an elusive response. In particular, this context-dependent behavior is modeled in the third step. More specifically, in the third step, the *Mental Integrity* vector and the contextual event label are mapped to a probability distribution function over the three response subsets, i.e. R_q^t , R_q^f and R_q^n , determining what type of response will be selected. For example, if the Virtual-Suspect is presented with a statement regarding the *Criminal* event and the Virtual-Suspect is mentally stable, i.e. $s' = (3, 0, 0)$, then there is a high probability of responding with a deceptive response. Specifically, the corresponding distribution probability function will be set to $p(r^* \in R_q^f) = 1$, $p(r^* \in R_q^t) = 0$ and $p(r^* \in R_q^n) = 0$. On the other hand, for the exact same statement, if the mental integrity is compromised, i.e. $s' = (0, 0, 3)$, indicating that the Virtual-Suspect’s cognitive capacity is diminished, then there is a low probability of responding with a deceptive response. Specifically, the corresponding distribution probability function will be set to $p(r^* \in R_q^f) = 0.1$, $p(r^* \in R_q^t) = 0.5$ and $p(r^* \in R_q^n) = 0.4$. Lastly, the associated responses subset is randomly selected using the corresponding distribution probability function and is denoted by R_q^* .

Finally, in the fourth step, the output response, r^* , is randomly selected from the previously selected subset, R_q^* , using a uniform distribution function, i.e. $r^* \sim U(R_q^*)$. For instance, if the deceptive responses subset was selected, i.e. $R_q^* = R_q^f$, and R_q^f contained five deceptive responses, then all five deceptive responses would have an equal chance to be selected as the final response of the Virtual-Suspect.

Experiment

An experiment was conducted in order to compare the *Virtual-Suspect Response Model* (RMVS) selection mechanism to a that of a human. The experiment was motivated by Lin et al. (Lin et al. 2014) study who showed that automated training agents with human-like behavior can im-

prove people performance. Participants were asked to read transcripts of three interrogation simulations and answer a series of questions. As described above, the interrogation scenario chosen for the experiment is based on an actual burglary case from early 2013 (see **Example 1**). In addition, the Virtual-Suspect’s personality profile was chosen to represent a moderately calm individual. More specifically, the initial internal-state vector was set to $s_0 = (0, 0, -3)$.

Prior to the experiment, three different Virtual-Suspect models were simulated. In the first, a human instructor acted as the suspect. The second was the Virtual-Suspect response model, i.e. RMVS. In the third, a baseline randomized response selection mechanism was used. In order to ensure a valid comparison, all three simulations’ responses were selected from the same response database. Three social sciences students aging from 23-26 were asked to participate in an interrogation simulation experiment. Their demographic characterization closely resemble those of police cadets. Each participant interacted with one of the above models; human instructor, VS agent and the random response model. The participants were presented with introductions explaining in detail the interrogation scenario, questioning techniques and rules of engagement with the suspect. The three participants were informed that the interrogation will terminate by eliciting a confession. However, if a confession could not be reached, the simulation will terminate after 30 minutes. Nevertheless, reaching a confession is rare and in all three simulations, the participants did not reach a confession. Fortunately, participants were able to present a high number of statements, ranging between 28-32 statements. In the experiment, 24 participants, ranging in age from 20-30, were asked to carefully read the three simulations’ transcripts and answer a series of questions. 12 of the participants were males and 12 were females. Again, the demographic of the participants closely resemble those of police cadets. The order in which the transcripts were presented was selected with great care in order to ensure that the results remain order independent. Consequently, each possible order of the three transcripts was presented to four participants, two males and two females.

Our hypothesis was that humans will find it difficult to differentiate between a human trainer and the Virtual-Suspect response model. However, the randomized response selection baseline mechanism will be distinguishable. The next section provides an in-depth presentation of the results following a discussion of their importance.

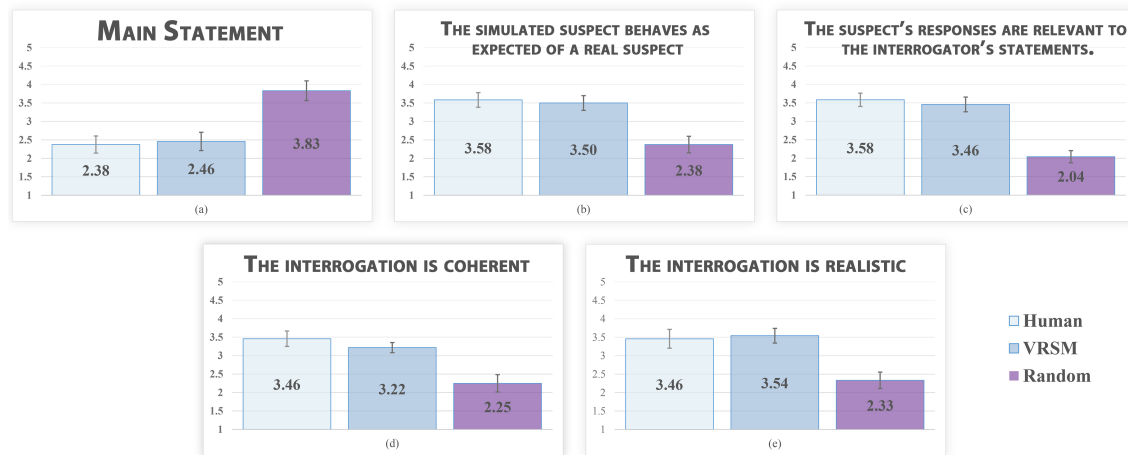


Figure 4: Experiment's Results

Results

As described above, participants were asked to read the three transcripts and answer a series of questions. Since the experiment's hypothesis was that humans will find it difficult to differentiate between a human and the RMVS, the participants were presented with the following question:

"The suspect's responses in the transcripts have been selected from an existing response database. In some of the transcripts, the responses have been selected by a human participant, and in the others by a computer. Would you say that in this current transcription the responses were chosen by the human participant or by the computer?"

The participants were asked to specify their belief on a scale ranging from one to five, one being that the suspect's responses have been selected by a human participant and five being that the suspect's responses have been selected by a computer. The results are presented in Figure 4 (a). The average scores are 2.38, 2.46 and 3.83 for the Human, RMVS and the Random baseline, respectively (lower is better, i.e. perceived as more human).

To analyze the results, repeated measures ANOVAs were conducted to compare the three conditions, i.e. Human, Random and RMVS. The resulting ANOVAs are $F(2, 22) = 8.15$, partial $\eta^2 = 0.57$, $p < 0.002$. As hypothesized, the RMVS were rated significantly better, i.e. more human like, than the Random baseline (using paired t-test $p < 0.01$). Similarly, the Human transcript was also rated significantly better than the Random baseline (using paired t-test $p < 0.01$). Furthermore, the Human transcript was rated slightly better than the RMVS. However, this small difference was not found to be statistically significant. To further support our hypothesis, the results of the following four statements are presented in Figure 4 (b) to (e).

The participants were asked to specify their level of agreement or disagreement on a standard five level Likert-type scale for these four statements. The scale consists of the following five levels: 1. Strongly disagree, 2. Disagree, 3. Nei-

ther agree nor disagree, 4. Agree and 5. Strongly agree. The averages and standard deviation results are presented in Figure 4 (b) to (e). Running repeated measures ANOVAs on the three conditions yielded similar results to the results for the main question. In addition, as can be seen in the figure, the RMVS performed significantly better than the Random baseline (using a paired t-test $p < 0.01$). Similarly, the Human transcript also performed significantly better than the Random baseline (using a paired t-test $p < 0.01$). Furthermore, the rates of Human and RMVS were similar with no statistically significant difference. To conclude, the results support the experiment's hypothesis. People were not able to distinguish between the Human suspect and the RMVS. However, they were able to easily distinguish between the Random baseline and the RMVS as well as between the Random baseline and the Human.

Conclusion

In this paper the Virtual-Suspect system was presented, which consists of a configurable simulation and the Virtual-Suspect Response Model (RMVS). The *Scenario Configuration* and *Interrogative Interviewing* databases enable the simultaneous simulation of multiple interrogation cases. Furthermore, the Virtual-Suspect's personality profile can be configured which is based on Eysenck's PEN Model. In addition, the Virtual-Suspect Response Model (RMVS) was presented. For every interrogator's statement, the RMVS updates the internal-state vector, converts it into the Mental-Integrity vector which determines the probability of responding with a truthful, deceptive or a neutral response. Lastly, a perceptual experiment was conducted comparing the RMVS, Human and Random responses. The results suggest that humans have difficulty differentiating between simulations generated by the RMVS and those of a human. We are in the process of integrating the VS system in the interrogating interview training program of a national police academy.

References

- Anderson, K.; André, E.; Baur, T.; Bernardini, S.; Chollet, M.; Chryssafidou, E.; Damian, I.; Ennis, C.; Egges, A.; Gebhard, P.; et al. 2013. The tardis framework: intelligent virtual agents for social coaching in job interviews. In *Advances in computer entertainment*. Springer. 476–491.
- Bruijnes, M.; Wapperom, S.; Heylen, D.; et al. 2014. A virtual suspect agents response model.
- Eysenck, H. J. 1990. *Biological dimensions of personality*. Guilford Press.
- Hubal, R. C., and Frank, G. A. 2001. Interactive training applications using responsive virtual human technology. *children* 21:25.
- Kenny, P.; Hartholt, A.; Gratch, J.; Swartout, W.; Traum, D.; Marsella, S.; and Piepol, D. 2007. Building interactive virtual humans for training environments. In *Proceedings of IITSEC*, volume 174.
- Kraus, S., and Lehmann, D. 1995. Designing and building a negotiating automated agent. *Computational Intelligence* 11(1):132–171.
- Lin, R., and Kraus, S. 2010. Can automated agents proficiently negotiate with humans? *Communications of the ACM* 53(1):78–88.
- Lin, R.; Gal, Y.; Kraus, S.; and Mazliah, Y. 2014. Training with automated agents improves people’s behavior in negotiation and coordination tasks. *Decision Support Systems* 60:1–9.
- Lin, R.; Oshrat, Y.; and Kraus, S. 2009. Investigating the benefits of automated negotiations in enhancing people’s negotiation skills. In *Proceedings of The 8th International Conference on Autonomous Agents and Multiagent Systems-Volume 1*, 345–352.
- Luciew, D.; Mulkern, J.; and Punako, R. 2011. Finding the truth: interview and interrogation training simulations. In *Proceedings of the Interservice/Industry Training, Simulation, and Education Conference (IITSEC)*.
- Olsen, D. E. 1997. Interview and interrogation training using a computer-simulated subject.
- Parsons, T. D.; Kenny, P.; Ntuen, C. A.; Pataki, C. S.; Pato, M. T.; Rizzo, A. A.; St-George, C.; and Sugar, J. 2008. Objective structured clinical interview training using a virtual human patient. In *Medicine Meets Virtual Reality 16-Parallel, Combinatorial, Convergent: NextMed by Design, MMVR 2008*.
- Roque, A., and Traum, D. 2007. A model of compliance and emotion for potentially adversarial dialogue agents. In *Proceedings of the 8th SIGdial Workshop on Discourse and Dialogue*, 35–38.