Tommaso Fornaciari

Deception Detection in criminal analysis: from lie detector to stylometry

PAN 2015
Tolouse, September 9th, 2015

Outline

Outline

Definition

Psychologist define deception in humans as *"an act that is intended to foster in another person a belief or understanding which the deceiver considers to be false"* (Zuckerman et al., 1981).

To detect deception means to detect the presence of a mental state in a person.

The problem is that we have no direct access to the mental states (Granhag et al., 2015).

Therefore, it is necessary to identify measurable signs, be they physiological or behavioral, associated with the mental state of interest.

The perfect cue of deception would appear if and only if the mental state is present: like the Pinocchio's growing nose. But... does it exists?

Human performance

Human performance in deception detection:

Human performance

Human performance in deception detection:

- is not better than chance (Bond and De Paulo, 2006).
  O'Sullivan and Ekman (2004) found that some people - they call *"wizards"* - are particularly skilled, but Bond and Uysal (2007) analysed their results and concluded that *"chance can explain results that the authors attribute to wizardry"*.

Human performance

Human performance in deception detection:

- is not better than chance (Bond and De Paulo, 2006).
  O'Sullivan and Ekman (2004) found that some people - they call *"wizards"* - are particularly skilled, but Bond and Uysal (2007) analysed their results and concluded that *"chance can explain results that the authors attribute to wizardry"*.

- does not improve even after specific training.
  Kassin and Fong (1999) tried to develop lie detection training procedures to be employed in forensic settings, however Levine et al. (2005) claim that this is not particularly effective to improve the ability of subjects.

Cues of deception

Based on the kind of clues of deception which are examined, in literature there can be found studies focusing on:

- Physiological variables:
- Non-Verbal behavior:
- Verbal behavior:

Cues of deception

Based on the kind of clues of deception which are examined, in literature there can be found studies focusing on:

- Physiological variables:
  - Complex recording tools; ☹
  - High subjects' cooperation; ☹
  - Objective cues' measurement. ☺
- Non-Verbal behavior:
- Verbal behavior:

Cues of deception

Based on the kind of clues of deception which are examined, in literature there can be found studies focusing on:

- Physiological variables:
  - Complex recording tools; ☹
  - High subjects' cooperation; ☹
  - Objective cues' measurement. ☺
- Non-Verbal behavior:
  - Simple to complex recording tools; 😐
  - Low to high subjects' cooperation; 😐
  - Subjective to objective cues' measurement; 😐
- Verbal behavior:

Cues of deception

Based on the kind of clues of deception which are examined, in literature there can be found studies focusing on:

- Physiological variables:
  - Complex recording tools;
  - High subjects' cooperation;
  - Objective cues' measurement.
- Non-Verbal behavior:
  - Simple to complex recording tools;
  - Low to high subjects' cooperation;
  - Subjective to objective cues' measurement;
- Verbal behavior:
  - Simple recording tools;
  - Low subjects' cooperation;
  - Subjective to objective cues' measurement;

Leakage

The leakage of cues of deception is commonly expected to be greater within the behavioral areas not under conscious control.

According to mostly unspoken beliefs, such areas would be characterised as follows:

Leakage

The leakage of cues of deception is commonly expected to be greater within the behavioral areas not under conscious control.

According to mostly unspoken beliefs, such areas would be characterised as follows:

- Physiological variables: No conscious control.
- Non-Verbal behavior: Partial conscious control;
- Verbal behavior: Full conscious control;

But… Is it true?

Data collection

Techniques of interview can be associated to every kind of experimental paradigm.

However, when the verbal behavior is the object of analysis, the techniques of interview are themselves the tool of data collection:

- Physiological variables $\longrightarrow$ Technical devices
  + techniques of interview;
- Non-verbal behavior $\longrightarrow$ Video/audio recording and/or human observation
  + techniques of interview;
- Verbal behavior $\longrightarrow$ Techniques of interview
  + audio recording.

Lab and field studies

Studies about deception are differentiated by their experimental design:

- Laboratory studies:
- Field studies:

Lab and field studies

Studies about deception are differentiated by their experimental design:

- Laboratory studies:
  - Control of the variables; ☺
  - Ground truth known; ☺
  - Possibility of replication experiments. ☺
  - Lack of ecological validity; ☹

- Field studies:

Lab and field studies

Studies about deception are differentiated by their experimental design:

- Laboratory studies:
    - Control of the variables; ☺
    - Ground truth known; ☺
    - Possibility of replication experiments. ☺
    - Lack of ecological validity; ☹

- Field studies:
    - No control of variables; ☹
    - Ground truth often unknown; ☹
    - No possibility of replication experiments. ☹
    - Ecological validity. ☺

Outline

Lie Detector

The polygraph, better known as Lie Detector, is a device that was realised in 1921 by John Augustus Larson (Britannica, 2003), and records some bodily activities:

- Electro-Dermal Activity (EDA);
- Blood pressure;
- Pulse;
- Respiration.

Lie Detector

In the United States the debate regarding admissibility in Court and effectiveness of physiological measures for the evaluation of truthfulness in testimonies began on 1923, in the famous case of Frye vs. United States (Saxe and Ben-Shakhar, 1999).

The polygraph does not 'read the mind', but simply measures physiological variables, which are assumed to be associated to deception.
This association can be of two different kinds, which lead to two different strategies in the use of polygraph:

- Concern approach;
- Orienting reflex approach.

Concern approach

Assumption Polygraph can be employed to detect signs of stress which are supposed to be related to the production of deceptive statements.

Protocol Control Question Test (CQT), an interview protocol in five phases aimed to check the bodily reactions of the subjects to crime-related and control questions (Reid, 1947; Raskin, 1986).

Performance Field studies showed that 83% to 89% of liars were correctly classified.
53% and 75% of innocent examinees were correctly identified, but a quota from 12% to 47% was misclassified (Vrij, 2008).

Pro/Cons High accuracy in detecting liars. ☺
Vulnerability to false-positive errors. ☹

Orienting reflex approach

Assumption *"An orienting response [...omissis...] occurs when someone is confronted with a personally significant stimulus"* (Vrij, 2008).

Protocol Concealed Information Test (CIT) (Verschuere et al., 2011), originally known as Guilty Knowledge Test (GKT), which implies the presentation of stimuli, usually images, the subjects should be familiar with.

Performance In field studies, the protocol achieves 76% to 88% of accuracy in identifying liars.
Only 1% to 6% of innocent subjects were incorrectly classified (Vrij, 2008).

Pro/Cons Resistance to false-positive errors.
Some weakness with false-negative errors.
Applicability limited to specific settings.

Event-Related Potentials

Electroencephalographic (EEG) brain event-related potentials (ERPs) can support the application of the GKT protocol.

While the Autonomic Nervous System (ANS) responses take several seconds to manifest, the P300 wave is much faster: its name comes from the typical peak latency of about 300 ms., which can actually be of 500-600 ms. for complex stimuli.

"P300 is sensitive to the rarity and meaningfulness of a stimulus, and the amplitude varies with the strength of recognition memory" (Granhag et al., 2015).

Features

P300-GKT is 100% accurate in detecting memories stored in the brain. ☺

Features

P300-GKT is 100% accurate in detecting memories stored in the brain. ☺

It is not possible to know which memory is crime-related. ☹

Features

P300-GKT is 100% accurate in detecting memories stored in the brain. ☺

It is not possible to know which memory is crime-related. ☹

P300 has low spatial and high temporal resolution. 😐

Features

P300-GKT is 100% accurate in detecting memories stored in the brain. ☺

It is not possible to know which memory is crime-related. ☹

P300 has low spatial and high temporal resolution. 😐

P300-GKT is rarely employed in police practice. ☹

Functional Magnetic Resonance Imaging

One of the most innovative approaches to deception deception relies on modern techniques of neuro-imaging.

The functional Magnetic Resonance Imaging (fMRI) detects activity of the brain areas, measuring changes in:

- Blood flow;
- Oxygen consumption.

Myth...

The use of fMRI to detect deception represents a fascinating perspective: in the United States some private laboratories already provide services of deception detection based on fMRI technology.

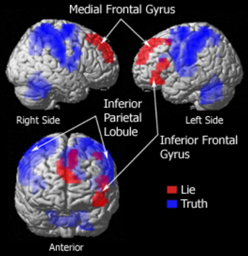


Image from noliemri.com

Myth...

The use of fMRI to detect deception represents a fascinating perspective: in the United States some private laboratories already provide services of deception detection based on fMRI technology.

However, Cohen (2012) claims that the advertisement of the technology *"violates consumer protection law under the Federal Trade Act."*

In fact, the findings in the scientific literature are not consolidated yet and only concern laboratory studies.

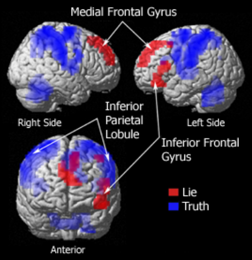


Image from noliemri.com

...and reality

The survey of the literature of Granhag et al. (2015) shows that 10 studies from 2005 and 2011 reached an overall accuracy from 69% to 100%. ☺

...and reality

The survey of the literature of Granhag et al. (2015) shows that 10 studies from 2005 and 2011 reached an overall accuracy from 69% to 100%. ☺

Experimental designs are far from realistic scenarios. *"Deception detection accuracy in the study with perhaps the most elaborate mock crime scenario used so far (Kozel et al., 2009) was rather low (67%)."*(Granhag et al., 2015) ☹

...and reality

The survey of the literature of Granhag et al. (2015) shows that 10 studies from 2005 and 2011 reached an overall accuracy from 69% to 100%. ☺

Experimental designs are far from realistic scenarios. *"Deception detection accuracy in the study with perhaps the most elaborate mock crime scenario used so far (Kozel et al., 2009) was rather low (67%)."*(Granhag et al., 2015) ☹

Ganis et al. (2011) found that, when the subjects were instructed to act countermeasures - mental actions carried out in front of irrelevant stimuli, in order to increase their saliency - the sensitivity (proportion of deceptive cases correctly classified) fell from 100% to 33%. ☹

Other features

Few replications are available. 😐

Other features

Few replications are available. 😐

fMRI allows a rich data collection. 🙂

Other features

Few replications are available. 😐

fMRI allows a rich data collection. 🙂

fMRI is characterised by high spatial and low temporal resolution. 😐

Outline

Methods

Studies focused on non-verbal clues of deception usually rely on the activity of trained raters who watch videos in which liars and true tellers interact, with the aim of analyzing some form of non-verbal behavior.

Coding systems are adopted in order to detect frequency, duration and intensity of several non-verbal cues and to compare the results for liars and true tellers.

Aim of these studies is to detect liars rather than single lies.

Assumptions

As formalised by Zuckerman et al. (1981), clues of deception might be found exploring the following factors:

Emotional reactions. Liars may experience feelings of guilt and fear of being unmasked, which could elicit anxiety signs.

Cognitive effort. Liars have to accomplish several tasks:

- To formulate narratives different from the truth;
- To be plausible and to not fall into contradiction;
- To check the others' reactions.

Liars are supposed to show more hesitations, more speech latencies and to reduce gestures of illustration.

Attempted behavioral control. Liars must be convincing. This task could be difficult, since some bodily reactions are almost beyond the voluntary control.
Discrepancies are expected between verbal and non-verbal behaviors, or different non-verbal behaviors.

Ekman's studies

The studies of Ekman rely on the idea that strong emotions can activate facial muscles almost automatically.

Two kind of signals which may be issued by the subjects:

Leakage cues. Behavioral expressions that the liars fail to squelch.

Deception cues. They share the same nature of the previous cues, but they are so brief that the emotion which caused them cannot be recognized.

Subjects can suppress their expressions within 1/25 sec, but this lapse of time is enough for a trained observer to detect such micro-expressions (Ekman, 2001).

Ekman's studies

The work of Ekman is famous, but questioned, especially for the inability to replicate facial coding:

"In brief, [...] the less aware we are of a behaviour, the more likely the behaviour is to signal a lie.
Subsequent scientific data have not been supportive of the original leakage theory, and subsequent work by Ekman and his colleagues shifted to focus more on the face and micro-expressions. Even with these modifications, however, leakage theory remains controversial" (Granhag et al., 2015; Weinberger, 2010)

Vrij's review

In a more recent literature review, Vrij (2008) took into consideration the study of De Paulo et al. (2003) and summarised a set of 132 studies focused on non-verbal cues to deception.

He selected a restricted number of cues, whose the analyses in literature he considered particularly reliable.

They were divided as follows:

- 7 Vocal cues;
- 10 Visual cues.

Vocal cues

1. Speech hesitations: use of speech fillers e.g., ‘ah’, ‘um’, ‘er’, ‘uh’ and ‘hmmm’;
2. Speech errors: grammatical errors, word and/or sentence repetition, false starts, sentence change, sentence incompletions, slips of the tongue, etc.;
3. Pitch of voice: changes in pitch of voice, such as rise in pitch or fall in pitch;
4. Speech rate: number of spoken words in a certain period of time;
5. Latency period: period of silence between question and answer;
6. Pause durations: length of silent periods during speech;
7. Frequency of pauses: frequency of silent periods during speech (Vrij, 2008).

Visual cues

1. Gaze: looking into the face of the conversation partner;
2. Smile: smiling and laughing;
3. Self-adaptors: scratching the head, wrists, etc.;
4. Illustrators: hand and arm movements designed to modify and/or supplement what is being said verbally;
5. Hand and finger movements: movements of hands or fingers without moving the arms;
6. Leg and foot movements: movements of legs and feet;
7. Trunk movements: movements of the trunk;
8. Head movements: head nods and head shakes;
9. Shifting position: movements made to change seating position;
10. Blinking: blinking of the eyes (Vrij, 2008).

Vrij's results

The effect sizes, evaluated by Vrij (2008), were found significant for only three cues:

1. Pitch: liars use a higher pitch of voice than truth tellers, but the effect is small.
   Furthermore, the difference between liars and true tellers usually is only few Hertz, and needs professional devices to be detected;
2. Illustrators: liars show fewer illustrators than true tellers, with a 'small' effect size;
3. Hand and finger movements: liars move hands and fingers less than true tellers, with a 'small/medium' effect size.
   However, Vrij et al. (1997) analyzed this variable on 181 subjects, finding that '64% of them showed a decrease in hand/finger movements during deception, whereas 36% showed an increase of these movements during deception'.

Vrij's results

The overall results *"show an erratic pattern and indicate that many conflicting results have been found"* (Vrij, 2008):

- In some studies speech hesitations are more frequent in liars than in true tellers, in others they are less frequent.
- The pauses in the speech seem to be longer in liars than in true tellers, but not necessarily more frequent.
- Gaze behavior does not seem to be related to deception, even though as popular opinion is that liars tend to look away from their interlocutor. However this behavior is easy to control and people are aware of its importance for communication, thence it cannot be considered an effective marker for deception (Vrij, 2008).

Discussion

In spite of the wide set of cues considered by De Paulo et al. (2003), most of them showed non-significant trends.

The outcome of different studies are often inconsistent.

The behavioral cues reveal emotions: they are not specific for deception.

There is evidence that clusters of cues can show patterns of deceit, even though their composition may change in different situations.

"A cue akin to Pinocchio's growing nose does not exist" (Vrij, 2008).

Outline

Levels of analysis

To deal with language means to face two orders of complexity:

Semantics. Semantic analyses are aimed to collect and analyse the information, through the examination of:

- The internal logic of the narrative, that is possible contradictions or discrepancies between statements;
- The external logic, which concerns the relation between statements and objective elements (Smirnov, 1988).

Stylistics. Stylistic analyses are aimed to evaluate the reliability/truthfulness of the narrative, through the examination of:

- The linguistic style of the narrative, that is its degree of similarity with stylistic models which supposedly distinguish truthful from deceptive communications.

Approaches

For the purposes of deception detection, semantic data and stylistic patterns are often jointly evaluated by trained experts.

- Statement Validity Assessment (SVA);
- Reality Monitoring (RM);

In police practices, however, techniques of interview were developed, not necessarily addressing deception detection, but specifically aimed to the collection of information. Different techniques were proposed for two main scenarios, that is the interview of cooperative or uncooperative subjects.

- Cognitive interview;
- Reid Technique.

In the last 10-15 years, in the field of Natural Language Processing (NLP), new approaches to deception detection arose, relying essentially on the analysis of stylistic features, mostly automatically collected.

Statement Validity Assessment

Statement Validity Assessment (SVA) is probably the most employed verbal veracity assessment tool in forensic practice, accepted as evidence in Courts in North America, Austria, Germany, Sweden, Switzerland, and the Netherlands (Vrij, 2008).

Developed by Undeutsch (1967), SVA was aimed to evaluate the reliability of the testimonies of children in cases of suspect sexual abuses.

The assumption is the so-called Undeutsch hypothesis (Steller, 1989): the cognitive elaboration of a memory differs from the elaboration of an imaginative construction, and this difference should be perceivable in the features of the narratives.

SVA's phases

SVA consists of four steps:

1. A preliminary analysis of the case;
2. A semi-structured interview aimed to get the statements of the subject;
3. The Criteria-Based Content Analysis (CBCA), the core of SVA;
4. An evaluation of CBCA through the Validity Checklist.

CBCA, in turn, consists of 19 criteria, marked as present or absent by trained evaluators.

The Validity Checklist addresses possible effects of intervening variables, such as psychological characteristics and motivation of the subject and of the interviewer.

CBCA's criteria

- General characteristics:
  1. Logical structure;
  2. Unstructured production;
  3. Quantity of details;
- Specific contents:
  1. Contextual embedding;
  2. Descriptions of interactions;
  3. Reproduction of conversation;
  4. Unexpected complications during the incident;
  5. Unusual details;
  6. Superfluous details;
  7. Accurately reported details misunderstood;
  8. Related external associations;
  9. Accounts of subjective mental state;
  10. Attribution of perpetrator's mental state;
- Motivation-related contents:
  1. Spontaneous corrections;
  2. Admitting lack of memory;
  3. Raising doubts about one's own testimony;
  4. Self-deprecation;
  5. Pardoning the perpetrator;
- Offence-specific elements:
  1. Details characteristic of the offense.

SVA's performance

Laboratory studies suggest that CBCA can identify truth and lies with a degree of accuracy of around 70% (Vrij, 2008).
Unfortunately, Undeutsch (1984) claimed that lab studies are not particularly useful in testing the SVA, as they lack of ecological validity.

By contrast, to evaluate field studies is often impossible as convictions and confessions, which should be used to establish the ground truth, frequently result from the employment of SVA itself.

Nevertheless, Vrij (2008) finds that one of the most reliable field studies shows *"several, albeit small, differences between truthful and fabricated statements (Lamb et al., 1997), and all of these differences were predicted by the Undeutsch hypothesis."*

Reality Monitoring

Reality Monitoring (RM), developed by Johnson and Raye (1981), relies on the idea that cognitive processes related to perceived and imagined events are different.

Similarly to SVA, the RM calls for checking the presence/absence of RM criteria in the subjects' statements.

RM is not widely employed in forensic practice, maybe because it does not address directly deception.

RM criteria:

1. Clarity;
2. Perceptual information;
3. Spatial information;
4. Temporal information;
5. Affect;
6. Reconstructability of the story;
7. Realism;
8. Cognitive operations.

RM and LIWC

Bond and Lee (2005), in order to verify the presence of the RM criteria, annotated the transcripts of their interviews both manually and using the Linguistic Inquiry and Word Count (LIWC) the well known lexicon created by Pennebaker et al. (2001). In the first case the authors found differences between liars and truth tellers, in the second one they did not.

The opinion of Vrij (2008) is that *"the problem with using automatic coding is that computer word counting systems ignore context, whereas the RM tool, as well as CBCA, require that the context is taken into account."*

In literature, the accuracy of RM in classifying the statements is remarkable: *68.8%* (Vrij, 2008).

Cognitive interview

Theoretically well-grounded on memory theory, the Cognitive Interview (CI) is a well-known protocol for testimonies' collection, widely employed in police investigations.

The tool, realised by Fisher and Geiselman (1992), is aimed to enhance the recollection of detailed information in cooperative eyewitnesses.

During the CI, are asked to carry out a number of task, such as:

- Mental reinstatement of environmental and personal contexts;
- In-depth reporting (possibly with 'think-aloud' technique);
- Reporting the event in different orders;
- Reporting the event from different perspectives.

There is experimental evidence that, compared to standard interview, CI increases the leakage of cues of deception as well (Colwell et al., 2002).

Reid technique

The Reid technique (Inbau et al., 2011) is a method for police interrogations, developed since 1947 and widely employed, especially in United States.

The basic idea is to put the uncooperative examinees under psychological pressure through the use of emotionally charged questions and argumentations, in order to elicit cues of deception and the confession of the crime.

The Reid techniques is criticised both for the weakness of the theoretical basis and for the risk of leading to false confessions (Gallini, 2010).

Nine steps

The Reid technique is a structured nine-steps process, which includes confrontation and minimalisation strategies (Moore and Fitzsimmons, 2011):

- At the beginning the suspect is informed there is evidence (either real or not) he is guilty: the subject is interrupted if he tries to deny the allegations.
- Moral justifications or rationalizations for the crime are presented, suggesting that the confession lead to leniency;
- The questions to the subject imply presumption of guilty;

"From the suspect's perspective, isolation, fatigue and fear may produce a compliant (but false) confession from a person who merely wants to extricate himself from an aversive situation and/or who succumbs to implied threats of dire consequences or implicit promises of clemency" (Moore and Fitzsimmons, 2011).

The case of Juan Rivera

"In 1993, Juan Rivera, a resident of Waukegan, Illinois, was sentenced to life in prison for the rape and murder, a year earlier, of an eleven-year-old girl".

"No physical evidence linked him to the attack".

"Nevertheless, in late October of 1992, he was brought to Lake County Jail, in Waukegan, and interrogated intermittently for four days. Twice during that time, Rivera was taken to Reid headquarters, in Chicago, where a Reid employee named Michael Masokas administered polygraph tests. The results were mixed, but Masokas told Rivera that the evidence demonstrated his guilt. Eventually, after more round-robin interrogation, he signed a confession.
Rivera's conviction was affirmed three times [...]. The last trial was, in many ways, the most astonishing, because it came four years after new DNA evidence had exculpated Rivera. Nevertheless, he was found guilty again, based partly on the strength of his original confession. Rivera's attorneys appealed, and he was released in 2012.".

"John E. Reid & Associates will pay two million dollars, which appears to be the largest settlement in its history" (Starr, 2015).

Outline

Stylometry and Deception Detection

Modern stylometry is the branch of Natural Language Processing (NLP) which studies texts considering their stylistic features only, making use of:

- Computational methods for automatic extraction of linguistic cues;
- Machine learning techniques for text analysis.

Stylometry has proven successful in detecting deception, dealing with:

- Spoken and written language in laboratory conditions (Newman et al., 2003; Strapparava and Mihalcea, 2009);
- Synchronous and asynchronous texts in Computer-Mediated Communication (Hancock et al., 2007; Zhou et al., 2004; Zhou, 2005);
- Spoken and written language collected on the field in judicial context (Bachenko et al., 2008; Fornaciari and Poesio, 2013).

Stylometry and Personality

Personality Recognition from texts is a computational linguistic task as well.

It consists in the automatic classification of authors' personality traits using textual cues as features.

Celli (2013) realised a semi-supervised system for personality recognition available online, which labels the texts with personality traits.

The study of Fornaciari, Celli and Poesio

In the study of Fornaciari et al. (2013), deception detection and personality evaluation were combined.

The typical task of text classification was employed to explore the relationship between personality types and linguistic style in deceptive communication, making use of:

- DeCour, a corpus of deceptive statements issued in high stakes conditions (Fornaciari and Poesio, 2012);
- The system of Celli (2013) for personality recognition.

In particular, models performance in detecting deception was compared with personality types of the subjects.

The corpus DeCour

DeCour - DEception in COURt - is a corpus constituted by the transcripts of 35 hearings:

- Issued by 31 subjects;
- Held in 4 Italian Courts: Bologna, Bolzano, Prato and Trento.

They come from criminal proceedings for calumny and false testimony, where:

- at least a *verbatim* transcription of a hearing exists, reporting statements of the defendant;
- the defendant is found guilty;
- a final judgment exists about the false testimony, which points out the lies told by the defendant.

Analysis units

The participants in the courtroom are:

- The examinee, who can have the status of witness or defendant;
- The Public Prosecutor;
- The defendant's lawyer;
- The Judge;
- Possibly, other expert witnesses.

The analysis units are the utterances issued by the examinee..

One or more consecutive utterances constitute a turn, and each turn is delimited by the intervention of other participants.

Annotation

The utterances were labeled as follows:

- **False.** Utterances which are clearly identified as false in the judgment or which, according to identified lies, seem to be false.
- **True.** Utterances consistent with the reconstruction of the facts.
- **Uncertain.** Utterances related to the facts under investigation, but of which the deceptiveness is not proved. Also utterances that logically cannot be either true or false, like questions or utterances stopped in mid-sentence.

DeCour's statistics

Label	Utterances	Tokens	
		with punct.	no punct.
False	945	15924	13376
True	1202	15456	12847
Uncertain	868	10439	8669
Total	3015	41819	34892

Three coders marked about 20% of DeCour. Kappa was used as metric for their agreement (Artstein and Poesio, 2008):

- Its value for the three classes was $k = .57$;
- The value for two classes - false vs. true and uncertain utterances - was $k = .64$: a moderate (Carletta, 1996) or substantial (Landis and Koch, 1977) agreement.

Personality

Personality is defined as an affect processing system that describes persistent human behavioural responses to broad classes of environmental stimuli (Adelstein et al., 2011) and characterizes a unique individual (Mairesse et al., 2007).

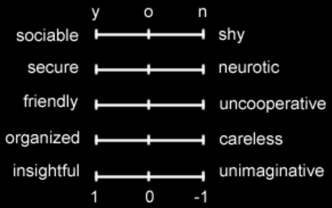
Differences in personality appear to include the style used in deception: e.g., several personality factors appear to correlate with the ability of a judge to detect deception (Enos et al., 2006).

Personality can be assessed by means of different questionnaires, such as the Big5 (Costa and MacCrae, 1992), that defines five bipolar traits and has become a standard over the years.

The dimensions of the Big5 were employed for the analyses of Celli (2013).

The Big Five personlity traits

Extraversion. It describes a person along the two opposite poles of sociability and shyness.



Emotional stability. Sometimes referred by its negative pole (neuroticism), Emotional stability describes the modality of impulse control along a scale that goes from control (a calm and stable person) to instability (an anxious and neurotic person).

Agreeableness. Agreeableness refers to the tendency to be sympathetic and cooperative towards others, rather than suspicious and antagonistic.

Conscientiousness. Conscientiousness describes a person in terms of self-discipline versus disorganization.

Openness. Openness to experience refers to the tendency to be creative and curious rather than unimaginative.

Features for personality recognition

As initial feature set, the system exploits language-independent correlations as features. These are taken from LIWC and MRC, whose correlations to personality are reported by Mairesse et al. (2007).

Feature	Ext.	Emo.	Agr.	Con.	Ope.
Punctuation	-.08**	-.04	-.01	-.04	-.1**
! marks	-.0	-.05*	.06**	.00	-.03
Word freq	.05*	-.06**	.03	.06**	-.07**
Numbers	-.03	.05*	-.03	-.02	-.06**
Parentheses	-.06**	.03	-.04*	-.01	-.1**
? marks	-.06**	-.05*	-.04	-.06**	.08**
Quotes	.05*	-.02	-.01	-.03	.09**
Repetitions	.05**	.1**	-.04*	-.05*	.09**

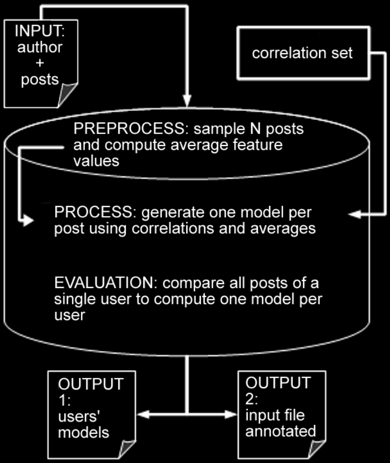
* $p < .05$
** $p < .01$

Pipeline for personality evaluation

Correlations are used as a model in a semisupervised way:

- in the preprocessing phase, labels are generated to retrieve label distribution;
- in the processing phase, labels are recomputed and filtered on the basis of the distribution found in the previous phase.

This system for personality recognition is the first to have been tested both on English and on Italian, obtaining F-measures between .63 and .68.



Experiments

We carried out three experiments:

Experiments

We carried out three experiments:

- In the first one, we simply used the five personality traits of the subjects as features for detecting deception at utterance level.
  We tested the performance over a number of different algorithms.
  In this experiment only true and false utterances were used, while the uncertain ones were not considered;

Experiments

We carried out three experiments:

- In the first one, we simply used the five personality traits of the subjects as features for detecting deception at utterance level.
  We tested the performance over a number of different algorithms.
  In this experiment only true and false utterances were used, while the uncertain ones were not considered;
- In the second experiment, we extracted surface features from our corpus in order to train models for the same task, in two different conditions:
  - to classify false vs. not-false utterances, that is vs. true and uncertain utterances together;
  - to classify false vs. true utterances.

  The methods of this experiment are described in the next slides;

Experiments

We carried out three experiments:

- In the first one, we simply used the five personality traits of the subjects as features for detecting deception at utterance level.
  We tested the performance over a number of different algorithms.
  In this experiment only true and false utterances were used, while the uncertain ones were not considered;
- In the second experiment, we extracted surface features from our corpus in order to train models for the same task, in two different conditions:
  - to classify false vs. not-false utterances, that is vs. true and uncertain utterances together;
  - to classify false vs. true utterances.

  The methods of this experiment are described in the next slides;
- In the end, we carried out a Multi-Dimensional Scaling - MDS Baayen (2008), in order to visualize the distances between our subjects, according to their personality traits. In particular, we used the MDS charts to compare the performance of the models with the different personality types of the subjects.

Training models

We trained models in order to classify the utterances of DeCour, according to the classes they belong to.

We tested a variety of classification methods, finding that the best performance was obtained with Support Vector Machines (SVMs) (Cortes and Vapnik, 1995).

Our SVM models were trained and then tested via *n*-fold cross-validations. In each experimental condition, the hearings of DeCour constitutes the folds for the cross-validations, so that the experiments were carried out with a 35-fold cross-validation.

The single utterances were our analysis units and they were described by vectors whose the values were the frequencies of *n*-grams of lemmas and part-of-speech (POS), collected separately form false and true utterances (the uncertain ones were not employed in the feature selection).

Feature selection and baselines

In order to select the features, we computed the Information Gain - IG (Forman, 2003) of the *n*-grams of lemmas and POS whose the frequency in DeCour was > 5, and we selected the *n*-grams having an IG value $> .01$.

The performance of the models was evaluated according to:

- majority baseline;
- a simple heuristic baseline:
  - The utterances beginning with the words Sì (Yes), Lo so (I know) and Mi ricordo (I remember) are classified as true;
  - The utterances beginning with the words No (No), Non lo so (I don't know) and Non mi ricordo (I don't remember) are classified as false;
  - All other utterances are randomly classified as true or false, according to the rate of true and false utterances present in DeCour.

Personality traits for deception detection

In the first experiment, we used personality traits as features for the classification task of the utterances in DeCour as true or false.

Algorithm	Acc.	M. Precision	M. Recall	F-measure
mbl (zeroR)	.5598	.313	.56	.402
dt (J4.8)	.5803	.579	.586	.55
nb (NaïveBayes)	.5631	.548	.562	.538
svm (SMO)	.5850	.582	.585	.533
ripper (JRip)	.5808	.576	.582	.532

Majority baseline: 55.98%
Heuristic baseline: 59.57%

Surface features for deception detection

False vs. Not-False utterances classification

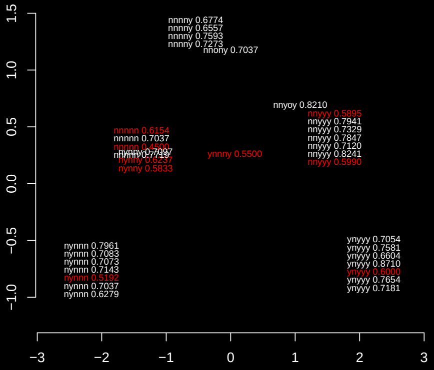
	Correctly classified entities	Incorrectly classified entities
False	342	284
Not-False	1786	603
Total accuracy	70.58%	29.42%
Majority baseline	68.66%	
Heuristic baseline	62.39%	

False vs. True utterances classification

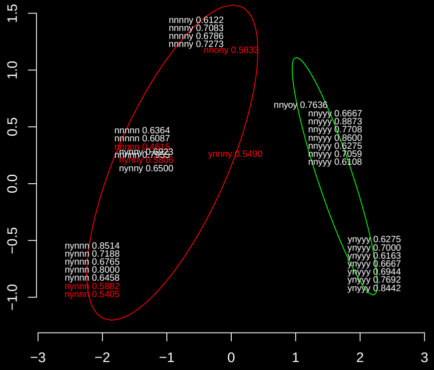
	Correctly classified entities	Incorrectly classified entities
False	511	234
True	968	434
Total accuracy	68.89%	31.11%
Majority baseline	55.98%	
Heuristic baseline	59.57%	

MDS - Personality and performance

False vs. Not-False utterances



False vs. True utterances



In red the performance below the heuristic baseline.

Personality and deception detection

In DeCour we have only 35 hearings and we found only 9 personality types, out of 120 possible profiles (five-factorial): the evaluation of the results must be prudent.

Personality traits are not particularly useful to detect deception at utterance level, since they supposedly identify the liars rather than the lies.

However:

- In the False vs. Not-False utterances experiment MDS seem to not show a clear pattern.
- In the False vs. True utterances experiment two clusters appeared. The T-Test confirmed that their accuracies belong to different populations.

Personality and models performance

The models performed better for subjects:

- extrovert, even though many subjects seemed to be introvert;
- friendly;
- organized;
- insightful, even though this trait belong to subjects difficult to be classified.

Instead, the models' performance was lower for subjects:

- uncooperative;
- secure. By contrast, most people seemed to be neurotic: this is probably due to the stress of the situation.

Outline

Cross-disciplinarity

The results of the studies on deception detection from different research field allow to draw some general conclusions:

- The perfect cue of deception, the Pinocchio's growing nose does not exist;
- A number of variables (most of them difficult to control in lab conditions) can affect the expression of possible cues of deception, so that in many cases the findings are inconsistent.
- Even in the few cases where the correlation between cue and deception is consistently found, the effect of the cues is weak;
- Deception detection can be improved by the evaluation of cluster of cues, even though their composition and their evaluation must take into account the operational context.

The role of DD

Every technique for detecting deception produces predictions whose the value is probabilistic.

The error rate is often remarkable, (sometime dramatic).

Therefore, the outcomes of Deception Detection analyses in criminal proceedings:

The role of DD

Every technique for detecting deception produces predictions whose the value is probabilistic.

The error rate is often remarkable, (sometime dramatic).

Therefore, the outcomes of Deception Detection analyses in criminal proceedings:

- Are not suitable for leading the final decision of the judge; ☹

The role of DD

Every technique for detecting deception produces predictions whose the value is probabilistic.

The error rate is often remarkable, (sometime dramatic).

Therefore, the outcomes of Deception Detection analyses in criminal proceedings:

- Are not suitable for leading the final decision of the judge; ☹
- Can represent a useful support for investigations: ☺

Future perspectives

Given the scarce reliability of the cues of deception, Vrij and Granhag (2012) suggest to focus not on finding out new cues - which is probably an approach doomed to fail - but on manipulating the interactions with the subjects, in order to enhance the expression of the deception clues already known.

Basically, two ways can be chosen:

Imposing emotional load. This is the path followed, for example, in the Reid technique (Inbau et al., 2011). The problem is that cues of emotions are not specific of deception.

Imposing cognitive load. By contrast, the idea of Vrij and Granhag (2012) is that increasing the cognitive load of the tasks does not affect remarkably the behavior of the truth-tellers, while enhances the leakage of cues of deception from liars: *"if lying requires more cognitive resources than truth telling, liars will have fewer cognitive resources left over."*

Computational linguistics

The pro and cons of computational linguistics in the field of deception detection may be summarised as follows:

Computational linguistics

The pro and cons of computational linguistics in the field of deception detection may be summarised as follows:

- The use of supervised methods imply a complex data collection for field application; ☹

Computational linguistics

The pro and cons of computational linguistics in the field of deception detection may be summarised as follows:

- The use of supervised methods imply a complex data collection for field application; ☹
- Limitations due to domain dependence must be considered; 😐

Computational linguistics

The pro and cons of computational linguistics in the field of deception detection may be summarised as follows:

- The use of supervised methods imply a complex data collection for field application;
- Limitations due to domain dependence must be considered;
- High granularity is allowed, as single utterances can be employed as analysis units;

Computational linguistics

The pro and cons of computational linguistics in the field of deception detection may be summarised as follows:

- The use of supervised methods imply a complex data collection for field application; ☹
- Limitations due to domain dependence must be considered; 😐
- High granularity is allowed, as single utterances can be employed as analysis units; ☺
- The data collection is not invasive; ☺

Computational linguistics

The pro and cons of computational linguistics in the field of deception detection may be summarised as follows:

- The use of supervised methods imply a complex data collection for field application; ☹
- Limitations due to domain dependence must be considered; 😐
- High granularity is allowed, as single utterances can be employed as analysis units; ☺
- The data collection is not invasive; ☺
- The methods do not require the cooperation of the subjects; ☺

Computational linguistics

The pro and cons of computational linguistics in the field of deception detection may be summarised as follows:

- The use of supervised methods imply a complex data collection for field application; ☹
- Limitations due to domain dependence must be considered; 😐
- High granularity is allowed, as single utterances can be employed as analysis units; ☺
- The data collection is not invasive; ☺
- The methods do not require the cooperation of the subjects; ☺
- The methods do not even require the interaction with the subjects; ☺

Computational linguistics

The pro and cons of computational linguistics in the field of deception detection may be summarised as follows:

- The use of supervised methods imply a complex data collection for field application; ☹
- Limitations due to domain dependence must be considered; 😐
- High granularity is allowed, as single utterances can be employed as analysis units; ☺
- The data collection is not invasive; ☺
- The methods do not require the cooperation of the subjects; ☺
- The methods do not even require the interaction with the subjects; ☺
- Linguistic analyses can be employed jointly with any paradigm of information assumption and with any other technology; ☺

Computational linguistics

The pro and cons of computational linguistics in the field of deception detection may be summarised as follows:

- The use of supervised methods imply a complex data collection for field application; ☹
- Limitations due to domain dependence must be considered; 😐
- High granularity is allowed, as single utterances can be employed as analysis units; ☺
- The data collection is not invasive; ☺
- The methods do not require the cooperation of the subjects; ☺
- The methods do not even require the interaction with the subjects; ☺
- Linguistic analyses can be employed jointly with any paradigm of information assumption and with any other technology; ☺
- The performance in detecting deception is similar to that of other methods. ☺

Thanks!

Gustav Klimt, *Nuda Veritas*, 1899
Österreichisches Theatermuseum - Vienna

Bibliography I

Adelstein, J. S., Shehzad, Z., Mennes, M., DeYoung, C. G., Zuo, X.-N., Kelly, C., Margulies, D. S., Bloomfield, A., Gray, J. R., Castellanos, F. X., et al. (2011). Personality is reflected in the brain's intrinsic functional architecture. *PloS one*, 6(11):e27633.

Artstein, R. and Poesio, M. (2008). Inter-coder agreement for computational linguistics. *Comput. Linguist.*, 34(4):555–596.

Baayen, R. (2008). *Analyzing linguistic data: a practical introduction to statistics using R.* Cambridge University Press.

Bachenko, J., Fitzpatrick, E., and Schonwetter, M. (2008). Verification and implementation of language-based deception indicators in civil and criminal narratives. In *Proceedings of the 22nd International Conference on Computational Linguistics - Volume 1*, COLING '08, pages 41–48, Stroudsburg, PA, USA. Association for Computational Linguistics.

Bond, C. F. and De Paulo, B. M. (2006). Accuracy of Deception Judgments. *Personality and Social Psychology Review*, 10(3):214–234.

Bond, C. F. and Uysal, A. (2007). On lie detection "wizard". *Law and Human Behavior*, 31(1):109–115.

Bibliography II

Bond, G. and Lee, A. (2005). Language of lies in prison: Linguistic classification of prisoners' truthful and deceptive natural language. *Applied Cognitive Psychology*, 19:313–329.

Britannica, E. (2003). Encyclopædia Britannica's Great Inventions.

Carletta, J. (1996). Assessing agreement on classification tasks: the kappa statistic. *Comput. Linguist.*, 22(2):249–254.

Celli, F. (2013). *Adaptive Personality Recogntion from Text.* Lambert Academic Publishing.

Cohen, J. T. (2012). Merchants of Deception: The Deceptive Advertising of fMRI Lie Detection Technology. *Seton Hall Legislative Journal*, 35(1):6.

Colwell, K., Hiscock, C. K., and Memon, A. (2002). Interviewing techniques and the assessment of statement credibility. *Applied Cognitive Psychology*, 16(3):287–300.

Cortes, C. and Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20.

Bibliography III

Costa, P. T. and MacCrae, R. R. (1992). *Revised NEO Personality Inventory (NEO PI-R) and NEO Five-Factor Inventory (NEO FFI): Professional Manual.* Psychological Assessment Resources.

De Paulo, B. M., Lindsay, J. J., Malone, B. E., Muhlenbruck, L., Charlton, K., and Cooper, H. (2003). Cues to deception. *Psychological Bulletin*, 129(1):74–118.

Ekman, P. (2001). *Telling lies: clues to deceit in the marketplace, politics, and marriage.* W.W. Norton.

Enos, F., Benus, S., Cautin, R. L., Graciarena, M., Hirschberg, J., and Shriberg, E. (2006). Personality factors in human deception detection: comparing human to machine performance. In *INTERSPEECH*, pages 813–816.

Fisher, R. P. and Geiselman, R. E. (1992). *Memory-enhancing techniques for investigative interviewing: The cognitive interview.* Charles C Thomas, Publisher.

Forman, G. (2003). An extensive empirical study of feature selection metrics for text classification. *J. Mach. Learn. Res.*, 3:1289–1305.

Bibliography IV

Fornaciari, T., Celli, F., and Poesio, M. (2013). The effect of personality type on deceptive communication style. In *European Intelligence and Security Informatics Conference, EISIC 2013*, pages 1–6, Uppsala, Sweden. IEEE Computer Society.

Fornaciari, T. and Poesio, M. (2012). DeCour: a corpus of DEceptive statements in Italian COURts. In Calzolari, N. C. C., Choukri, K., Declerck, T., Uäÿur Döäÿn, M., Maegaard, B., Mariani, J., Odijk, J., and Piperidis, S., editors, *Proceedings of the Eigth International Conference on Language Resources and Evaluation (LREC'12)*, Istanbul, Turkey. European Language Resources Association (ELRA).

Fornaciari, T. and Poesio, M. (2013). Automatic deception detection in italian court cases. *Artificial Intelligence and Law*, pages 1–38.

Gallini, B. (2010). Police'science'in the interrogation room: Seventy years of pseudo-psychological interrogation methods to obtain inadmissible confessions. *Hastings Law Journal*, 61:529.

Bibliography V

Gamer, M., Rill, H., Vossel, G., and Gödert, H. W. (2006). Psychophysiological and vocal measures in the detection of guilty knowledge. *International Journal of Psychophysiology*, 60:76–87.

Ganis, G., Rosenfeld, J. P., Meixner, J., Kievit, R. A., and Schendan, H. E. (2011). Lying in the scanner: covert countermeasures disrupt deception detection by functional magnetic resonance imaging. *Neuroimage*, 55(1):312–319.

Granhag, P. A., Vrij, A., and Verschuere, B. (2015). *Detecting Deception: Current Challenges and Cognitive Approaches.* John Wiley & Sons.

Hancock, J. T., Curry, L. E., Goorha, S., and Woodworth, M. (2007). On lying and being lied to: A linguistic analysis of deception in computer-mediated communication. *Discourse Processes*, 45(1):1–23.

Inbau, F., Reid, J., Buckley, J., and Jayne, B. (2011). *Criminal interrogation and confessions.* Jones & Bartlett Publishers.

Johnson, M. and Raye, C. (1981). Reality monitoring. *Psychological Review*, 88:67–85.

Bibliography VI

Kassin, S. and Fong, C. (1999). "i'minnocent!": Effects of training on judgments of truth and deception in the interrogation room. *Law and Human Behavior*, 23:499–516.

Kozel, A. F., Johnson, K. A., Grenesko, E. L., Laken, S. J., Kose, S., Lu, X., Pollina, D., Ryan, A., and George, M. S. (2009). Functional mri detection of deception after committing a mock sabotage crime*. *Journal of Forensic Sciences*, 54(1):220–231.

Lamb, M., Sternberg, K., Esplin, P., Hershkowitz, I., Orbach, Y., and Hovav, M. (1997). Criterion-based content analysis: A field validation study. *Child Abuse and Neglect*, 21:255–264.

Landis, R. J. and Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1):159–174.

Levine, T. R., Feeley, T. H., McCornack, S. A., Hughes, M., and Harms, C. M. (2005). Testing the Effects of Nonverbal Behavior Training on Accuracy in Deception Detection with the Inclusion of a Bogus Training Control Group. *Western Journal of Communication*, 69(3):203–217.

Bibliography VII

Mairesse, F., Walker, M. A., Mehl, M. R., and Moore, R. K. (2007). Using linguistic cues for the automatic recognition of personality in conversation and text. *J. Artif. Intell. Res.(JAIR)*, 30:457–500.

Moore, T. E. and Fitzsimmons, C. L. (2011). Justice imperiled: False confessions and the reid technique. *Crim. LQ*, 57:509.

Newman, M. L., Pennebaker, J. W., Berry, D. S., and Richards, J. M. (2003). Lying Words: Predicting Deception From Linguistic Styles. *Personality and Social Psychology Bulletin*, 29(5):665–675.

O'Sullivan, M. and Ekman, P. (2004). The wizards of deception detection. In Granhad, P. and Stromwall, L., editors, *Deception detection in forensic context*, pages 269–286. Cambridge Press, Cambridge, UK.

Pavlidis, J., Eberhardt, N., and Levine, J. (2002). Human behaviour: Seeing through the face of deception. *Nature*, 415:35–35.

Pennebaker, J. W., Francis, M. E., and Booth, R. J. (2001). *Linguistic Inquiry and Word Count (LIWC): LIWC2001*. Lawrence Erlbaum Associates, Mahwah.

Pepe, G., editor (1996). *La falsa donazione di Costantino*. Tea storica. TEA.

Bibliography VIII

Raskin, D. C. (1986). The polygraph in 1986: Scientific, professional, and legal issues surrounding acceptance of polygraph evidence. *Utah Law Review*, 29:29–74.

Reid, J. (1947). A revised questioning technique in lie detection tests. *Journal of Criminal Law, Criminology, and Police Science*, 37:542–547.

Sapir, A. (2000). *The LSI course on scientific content analysis (SCAN).* Laboratory for Scientific Interrogation, Phoenix, AZ.

Saxe, L. and Ben-Shakhar, G. (1999). Admissibility of polygraph tests: The application of scientific standards post- daubert . *Psychology, Public Policy, and Law*, 5(1):203 – 223.

Segrave, K. (2003). *Lie detectors: A social history.* McFarland.

Smirnov, V. (1988). Internal and external logic. *Bulletin of the Section of Logic*, 17(3):170–181.

Starr, D. (2015). Juan rivera and the dangers of coercive interrogation. *The New Yorker.*

Steller, M. (1989). Recent developments in statement analysis. In Yuille, J., editor, *Credibility Assessment*, pages 135–154. Kluwer, Deventer, The Netherlands.

Strapparava, C. and Mihalcea, R. (2009). The Lie Detector: Explorations in the Automatic Recognition of Deceptive Language. In *Proceeding ACLShort '09 - Proceedings of the ACL-IJCNLP 2009 Conference Short Papers*.

Undeutsch, U. (1967). Beurteilung der Glaubhaftigkeit von Aussagen [Veracity assessment of statements]. In Undeutsch, U., editor, *Handbuch der Psychologie: Vol. 11. Forensische Psychologie*, pages 26–181. Hogrefe, Gottingen, Germany.

Undeutsch, U. (1984). Courtroom evaluation of eyewitness testimony. *Applied Psychology*, 33(1):51–66.

Verschuere, B., Ben-Shakhar, G., and Meijer, E. (2011). *Memory detection: Theory and application of the Concealed Information Test*. Cambridge University Press.

Vrij, A. (2008). *Detecting Lies and Deceit: Pitfalls and Opportunities*. Wiley Series in Psychology of Crime, Policing and Law. John Wiley & Sons, 2nd edition.

Vrij, A. and Granhag, P. A. (2012). Eliciting cues to deception and truth: What matters are the questions asked. *Journal of Applied Research in Memory and Cognition*, 1(2):110–117.

Vrij, A., Winkel, F., and Akehurst, L. (1997). Police officers' incorrect beliefs about nonverbal indicators of deception and its consequences. In Nijboer, J. and Reijntjes, J., editors, *Proceedings of the first world conference on new trends in criminal investigation and evidence*, pages 221–238. Koninklijke Vermande, Lelystad, the Netherlands.

Weinberger, S. (2010). Airport security: intent to deceive? *Nature*, 465(7297):412–415.

Zhou, L. (2005). An Empirical Investigation of Deception Behavior in Instant Messaging. *IEEE Transactions on Professional Communication*, 48(2):147–160.

Zhou, L., Burgoon, J. K., Nunamaker, J. F., and Twitchell, D. (2004). Automating Linguistics-Based Cues for Detecting Deception in Text-based Asynchronous Computer-Mediated Communication. *Group Decision and Negotiation*, 13(1):81–106.

Bibliography XI

Zuckerman, M., DePaulo, B. M., and Rosenthal, R. (1981). Verbal and nonverbal communication of deception. *Advances in experimental social psychology*, 14(1):59.