detecting deception from speech content.

criteria-based content analysis.

In Germany in the 1950s the Supreme Court was concerned about relying on information provided (solely) by one or more young children to convict someone of the very serious crime of child sexual abuse. Therefore, it endorsed the idea that relevant experts (who are court appointed in the German inquisitorial criminal justice system) analyse such children’s accounts for indications that the children may well be describing genuinely experienced events. This analysis is usually referred to as criteria-based content analysis (CBCA). This analysis is part of ‘Statement Validity Assessment’, which courts in Germany, Sweden and the Netherlands have been using to guide their decisions. CBCA is based on a number of sensible assumptions, among which are that statements (i.e. the contents of what a person says) derived from memory of actual experiences differ in quality/content from those based on fabrication. Nineteen different criteria can be used to analyse the statements to help decide which are true. These criteria relate to (i) general characteristics of the statement (e.g. the amount of detail), (ii) specific contexts (e.g. reproduction of conversation, unexpected complications during the incident), (iii) motivation related contents (e.g. the child spontaneously correcting herself when giving her statement – something which liars may worry about doing) and (iv) details characteristic of that type of offence (e.g. that the child was ‘groomed’ before being abused).

However, it was not until many years after this procedure had been used in courts that research was conducted on its accuracy (for more on what constitutes accuracy see this chapter’s later section, ‘Can a testing procedure be relied upon?’). At a conference in Sweden in the early 1980s one of the ‘founding fathers’ of SVA/CBCA (Professor Udo Undeutsch) gave an Invited Lecture on this procedure which, for those invited delegates from the UK and the USA, was the first time they had heard in detail about this procedure and that it had already played a role in thousands of German court cases. At the end of his lecture I asked Professor Undeutsch what research had been conducted to assess its accuracy. He said that there had been none, the main reason for this being that, in his opinion, no research study could contain an event and subsequent interviewing that would approximate sexual abuse.

Nevertheless, in the last twenty years over thirty studies have examined CBCA and the general conclusion from them is that such analysis can discriminate between true and untrue accounts at a level above chance (averaging around seventy-two per cent), but not close to perfection. However, the various studies have found different (of the nineteen) criteria to discriminate between truthful and not truthful accounts. Some of these studies have analysed statements not of children (for whom the procedure was originally devised) but of adults (whose maturity could allow them purposely to provide some of the criteria in their false statements). Also, almost all of the studies have not been of real-life (see this chapter’s later section on the difficulty in real-life studies of establishing the ‘ground’ truth, that is whether a statement actually is true). Furthermore, to properly analyse statements using CBCA probably requires a lot of training, which some of the studies did not adequately provide. Even so, the general idea that a reasonable proportion of true statements differ from false statements in terms of their contents may have some merit when examining the statements of liars who are unaware of this idea.

Another approach, called ‘reality monitoring’, examines the content of what people say.

reality monitoring.

This approach to detecting deception is based on the assumption that memories based on experienced events (i.e. external sources) can be differentiated from memories based on imagining, thinking and reasoning (i.e. internal sources). That is, memories for what actually happened (truths) are different to some extent from made up stories (lies). Crucial to this approach is the notion that perceptual processes are very much involved in putting into memory truly experienced events involving information of a contextual (space, time), sensory (shapes, colours) and auditory (speech) nature. Thus true memories should contain such types of information, whereas lies involve many more thought processes (called cognitive operations).

A recent review of all previous studies found that the average accuracy of detecting truth/lies using this reality monitoring (RM) approach was around seventy per cent (when chance is fifty per cent). However, different studies had found different aspects of the RM criteria (i.e. information types) to be the most useful, and have defined the criteria in different ways. Furthermore, the delay between the event and the describing of it seems to affect the extent to which memories based on external sources (i.e. truths) still contain more contextual, sensory and semantic information. All of the studies so far published have been experiments con- ducted for research purposes (e.g. people lied/told the truth at the request of the experimenter). Real-life field studies are needed.

computer analysis.

A recent development in the analysis of speech to detect deception involves the use of computer software to analyse written transcripts of what people say. (These transcripts are written by humans who listen to tape recordings of people lying and telling the truth.) The software allocates each word spoken to a category (e.g. spatial, affective, cognitive) that theoretically may relate to lying/truth-telling in a way similar to the reality monitoring approach (see above). The software can also allocate the words to linguistic categories such as ‘negative emotions’ and ‘first person singular’ (e.g. I, me, my). However, at present the software has quite a high error rate (of around twenty per cent – which is not better than trained humans). Nevertheless, a recent study of prisoners lying and telling the truth about what happened in video clips that they had just seen found that these types of automatic computer-based transcript analyses to have a truth/lie detection rate significantly better than chance. However, a few word categories occurred more frequently in the way opposite to that predicted (e.g. more spatial words while lying). Clearly, more research is needed.

A recent study in the USA focused not on direct speech (e.g. one person speaking to another) but on the language used in messages such as e-mails. This pioneering study found that some of the cues were effective (e.g. deceivers displayed less ‘lexical diversity’, ‘content diversity’, and more ‘modifiers’). Contrary to previous research on direct speech, they found that in e-mail messages it was the deceivers who used more words (especially verbs, noun phrases and sentences). Again, much more research on this new topic is needed. (For more on forensic linguistics see chapter 7.)

combining behavioural and speech cues.

One possible way to improve lie detection is to combine the cues that have been found to be better than useless. If one analysed video tapes of people when known to be lying and telling the truth, one could (as described above) discover which cues (at least in those tapes, of those people, in that setting) occurred more (or less) often during the lies than during the truths. One could then analyse those video tapes (using only the valid cues) to see what success rate could be achieved. Professor Vrij did this using video recordings from two of our earlier studies involving nurses and students lying about a recent event.

In one of those studies liars showed fewer illustrators and hand/finger movements, longer response latency and more speech errors/hesitations, they also had a lower total CBCA score and RM score. (However, we must not forget that other studies have not found these behavioural cues to be associated with lying, but have found other cues to discriminate to a certain extent between truth and lies.) When all the cues found in the two studies to discriminate to some extent between truth and lies were combined, the resulting (complex statistical) analysis produced an accuracy rate of eighty-one per cent for the first study, eighty-eight per cent for the second study, and seventy-nine per cent for both studies’ data combined. (Note that these percentages are not based on humans making lie detection judgements but on counting up the cues, via video analysis, when the people were (i) lying and (ii) telling the truth.) These percent- ages indicate what the maximum possible accuracy rate should be if observers reliably used only the cues found in these studies actually to discriminate between lies and truth or (b) if suitable technology could ever be developed to monitor and quantify these cues.

Of course, other people (or the same people) in other situations might show different cues to lying. It is also very important to note that most of the cues employed in this combined approach relate to what people say and how they say it. If police interviews with possible liars are not conducted in a way that results in good samples of speech from the interviewees (see chapter 4), then a combined approach will be able to add little to the much lower detection rates that are typically found from the visual analysis of behaviour.

the polygraph.

You may have seen the movie Meet the Parents, in which a poly- graph test was used in attempting to detect deception.

The set of equipment known as the polygraph (from the Greek pol ’ = ‘many’ and graph = ‘to write’) measures various sorts of internal bodily activities such as heart rate, blood pressure, respiration and palmar sweating. These activities are displayed on charts or on computer screens. Such equipment is used in many medical and scientific settings. Its use in attempting to detect deception is based on the age-old assumption that lying is accompanied by changes in such internal bodily activities. While the equipment does measure such activities with great accuracy, the big issues for polygraphic lie detection are (i) whether deceivers’ bodily activities are reliably different when lying than when truth- telling and (ii) whether such differences do not similarly occur in truth-tellers (e.g. an innocent man being questioned about the murder of his wife with whom he was experiencing severe marital difficulties).

This section on polygraphic lie detection tests will focus on the following topics:

how to determine if a testing procedure can be relied upon;

its use in criminal investigations;

its use in security screening.

can a testing procedure be relied upon?

Psychologists around the world have devoted decades of effort to (a) establishing and publicizing how best to determine if tests can be relied upon and (b) assessing the quality of many thousands of tests. This is vitally important work. Just because a person or organization claims to have developed a useful test does not mean that the test is a good one. Many issues are relevant but the most important ones are reliability and validity.

Within psychology reliability refers not to accuracy but to similarity across time or among testers. It is the issue of validity that is closest to accuracy.

There are several aspects of reliability:

‘inter-examiner’ reliability which focuses on whether different testers make similar judgements to each other when assessing the same person

‘test-retest’ reliability which focuses on whether when re-tested a person receives a similar judgement as when first tested

‘inter-item’ reliability focuses on, for example, whether the various questions put to the person taking the test lead to the same conclusion.

Validity is concerned with the extent to which a test assesses what it claims to assess. It too has several aspects:

face validity is the extent to which a test (on the face of it) looks like it assesses what it claims.

content validity is concerned with the relationship of the con- tents of the test to the phenomenon being assessed.

construct validity concerns the relationship of the test to underlying theories/constructs concerning the phenomenon.

criterion validity is the extent to which scores on the test actu- ally predict outcomes (e.g. how accurate the procedure is at classifying people as lying or truth-telling).

incremental validity concerns how well a test compares with other tests that have been designed to examine the same phe- nomenon (e.g. detecting deceit).

All of the above forms of reliability and validity are crucially important in determining whether a testing procedure (such as polygraphic lie detection) actually is effective. Let us now look at research on whether polygraphic lie detection has been found to be accurate in the criminal setting.