the use of the polygraph in criminal investigations.

In the early days of attempting to detect criminal lying with a procedure involving the polygraph, a number of questioning techniques were developed, some of which have now been largely abandoned. For example, the relevant-irrelevant technique which compares physiological reactions to questions relevant to the crime with those not relevant. One of the major problems with this technique was that some innocent people (e.g. a loving husband) reacted strongest to relevant questions (e.g. ‘Did you murder your wife?’). Nowadays, one of two questioning techniques are usually employed, these being the Control Question Test and the Guilty Knowledge Test.

The Control Question Test (CQT) compares reactions to questions about the topic being investigated with reactions to questions that are thought to be arousing and that both guilty and innocent people will lie to. This procedure is based on the assumption that for an innocent person the latter type of questions will cause the greater reactions but that for a guilty person (who will be lying to all questions) the former type of questions will occasion larger reactions. Among the problems with the CQT is that innocent people may still react more strongly to the crime-relevant questions, especially if they fear that their answers to these will not be believed. Also, innocent people taking a polygraph test would be aware that if reliable information sup- porting their innocence were available, they would not need to take a polygraph test. Therefore, they know that it is difficult to prove their innocence and thus they are very concerned about the crime-relevant questions. This is probably why the CQT (for more on this see below) is not that good at correctly classifying innocent persons.

The Guilty Knowledge Test (GKT) compares reactions to various items, some of which may reveal knowledge of the crime. For example, the polygraph examiner may show a murder suspect several types of handgun, one of which is identical to the one used in the shooting. The suspect will be asked separately for each gun whether he or she recognizes it (or have ever touched it). This procedure is based on the assumption that the polygraphic reactions for guilty persons will be greater for the gun used in the shooting. So long as the alternatives (e.g. the various types of gun) are sufficiently similar to each other so that the ‘correct’ one does not stand out to an innocent suspect and the media or the police have not let slip (to the innocent suspect) what the ‘correct’ alternative is, then the GKT may not suffer as a procedure from as many problems at the CQT. However, as we shall see when we overview research studies of the GKT and the CQT, the former does at times seem to fail to correctly classify guilty people.

Most of the published research on the accuracy of use of the polygraph to detect deception has been focused on the criminal setting, but the vast majority of this research has not involved real crimes or real criminals. This is for a variety of reasons, including knowing the ‘ground truth’.

Ground truth involves knowing (independently, of course, of the polygraph testing procedure) whether the person being tested is lying or not. In real life (e.g. in police investigations) it is extremely difficult to be certain what took place during a crime and who was involved. A police suspect may be happy to admit that he or she was present but deny wrongdoing (e.g. state that sexual intercourse occurred by consent not rape). Unless there is rock-solid evidence that the person being tested is lying or truth- telling, the validity of the polygraph procedure cannot be properly assessed. Of course, if there is such rock-solid evidence, a poly- graph lie-detection test may well be superfluous.

Because ground truth is so difficult to assess in real-life investigations, the majority of the published studies that have tried to assess the effectiveness of the polygraph in criminal investigations have had to employ ‘mock’ crime scenarios.

laboratory studies.

A recent review of the possible effectiveness of the polygraph in mock crimes in the laboratory settings was provided by Professor Charles Honts. He noted for the CQT that once ‘inconclusive’ decisions had been removed from the data set the accuracy rate for ‘guilty’ persons (i.e. liars) was ninety-one per cent and for ‘innocent’ persons eighty-nine per cent. However, even though some researchers have shown ingenuity in designing their laboratory studies to have some aspects similar to real-life criminal investigations, for ethical (and other) reasons it is probably the case that laboratory studies will never be very similar to real life (e.g. the fear experienced by an innocent person accused of murder who knows he had a motive and was alone at the time of the killing).

field studies.

Because ground truth is so difficult to determine, relatively few ‘field’ (i.e. ‘real life’) studies of the effectiveness of polygraphic lie detection have been published. The 2004 report of the British Psychological Society (BPS) overviewed previously published reviews of such field studies, noting that determining ground truth via confessions (for example) is problematic because such confessions may have been, in part, affected by polygraph outcomes (e.g. a guilty person who passes the test may decide not to confess).

The majority of these field studies have employed the CQT – explained above. The BPS report (2004) noted on page 15 that ‘There is reasonable agreement between the reviews regarding guilty suspects. Correct classifications were made in 83 per cent to 89 per cent of the cases, whereas incorrect decisions (classifying a guilty suspect as innocent) were made in 10 per cent to 17 per cent of the cases.’ However, with regard to innocent suspects the report noted that these ‘are less encouraging. ... Depending on the review, between fifty-three per cent and seventy-eight per cent of innocent suspects were correctly classified and between eleven per cent and forty-seven per cent were incorrectly classified’ (p.15). (The review with the lowest rate of incorrect classifications of innocent suspects had the highest ‘inconclusive’ rate of twenty-nine per cent). Thus, as with laboratory mock studies using the CQT, while the error rate for incorrectly classifying guilty people is low, that for incorrectly classifying innocent people (i.e. as liars) is higher.

A few field studies have employed the Guilty Knowledge Test (GKT). The BPS report noted that these found high accuracy for classifying innocent suspects (ninety-four to ninety-eight per cent) but rather low accuracy for guilty suspects (forty-two to seventy-six per cent). This mirrors the outcomes of laboratory/ mock studies using the GKT. While its error rate for incorrectly classifying innocent people is low, that for incorrectly classifying guilty people (i.e. failing to detect liars) is higher.

Thus, the test/approach that has the stronger theoretical basis (the GKT) is the one that seems poorest at detecting the very people it was designed to detect (i.e. those with guilty knowledge). One reason for this is that guilty people, rather than innocent people, will be motivated to beat the test by the employment of what is referred to as countermeasures. These deliberate attempts may involve:

attempting to lessen physiological activity (i.e. what the poly- graph measures) in response to relevant questions (CQT) or items (GKT).

attempting to increase such activity in response to irrelevant questions or items.

The latter usually is easier to achieve than the former, unless one is well trained.

Many professional polygraphers claim that they believe they can detect the use of countermeasures. However, the quality published studies on this have shown that the use of counter- measures can be very effective.

use of the polygraph in security screening.

Another problem with regard to the test error rates (mentioned above) relates to the number of innocent suspects caught up in an investigation. Some countries are believed to employ polygraph testing to help determine who within, or wishing to join, their security services is a threat (e.g. a spy or a terrorist). They do this by testing a large number of people among whom few, if any, are a serious threat. In the USA the National Research Council (2003) pointed out that if, for example, the frequency of serious threat in an organization such as the security services is ten in ten thousand and a test procedure needs to detect at least eighty per cent of these threats, then over 1,600 people would fail the test. On the other hand, given that it is usually a considerable problem to misclassify innocent people (e.g. worthy employees) as threats, to have a much lower number of ‘false alarms’ (say about forty) would require setting a high degree of difference between truth-telling physiological activity and lying physiological activity which would result in eight of the ten threats passing the test. The inevitable weaknesses that exist in polygraphic lie detection led the National Research Council and the British Psychological Society to review the effectiveness of other possible methods to detect deception.

We mentioned above that in some countries the polygraph is used to test whether people who wish or are recommended to join the security services are telling the truth when being questioned about their intentions and their past. In a few countries some non- security organizations still, probably mistakenly, use polygraph testing to select employees even though after a review of the relevant research and human rights arguments the government of the USA in 1988 brought in the Polygraph Protection Act which prohibits the use of polygraph testing for employee selection (except by some government security agencies and in some strategic industries such as nuclear power).

Given that use of polygraph testing in pre-employment screening is beset by a number of problems, what can psychology offer? The BPS report overviews some other ways of assessing people’s honesty and integrity, for example voice stress analysis.

voice stress analysis.

In this chapter’s earlier section on observing behaviour we noted that some research studies have found that certain aspects of speaking (e.g. increases in voice pitch and speech errors) may be related (in some people, in some situations) to lying. While such increases in voice pitch may be small and hard to detect with the human ear, equipment has been designed that can accurately note such changes in pitch. Such equipment, often referred to as ‘voice stress analysers’, has been used in attempts to detect lying (though few studies on this have been published). For example, some insurance companies have been rumoured to be using voice stress analysers when people telephone to make a claim (e.g. that items were stolen while they were on holiday). The problems limiting the accuracy of such procedures are very similar to those for poly- graph testing (e.g. truthful people may be stressed and therefore their voice pitch rises and skilled liars may not be stressed).

In its overview the National Research Council (2003) concluded that ‘The practical performance of voice stress analysis for detecting deception has not been impressive’ (p.168) and that there is ‘little or no scientific basis for the use of ... voice measurement instruments as an alternative to the polygraph’ (p.168). However, whereas the polygraph usually requires the person being tested to be connected to the apparatus, voice stress analysis does not. This might result in fewer innocent/truthful people being stressed by the procedure, but it does raise issues of human rights (e.g. being aware that one is being tested). (Some new types of polygraph testing which monitor physiological activity without the person being aware of it may be under development.)

Conclusions.

Each of the methods and procedures used to try to detect deception has its severe limitations, many of which are inherent and therefore can never be overcome. Each seems to have error rates that are far from negligible. Even with training and experience (plus the most up-to-date equipment) people are typically far from perfect at catching liars. There is probably a very good reason for this. Humans have spent many thousands of years learning to deceive others. It would be naive of us, therefore, to think that lie detection will ever achieve very high accuracy rates. The 2004 BPS report said, ‘We must not deceive ourselves into thinking that there will ever be an error-free way of detecting deception’ (p. 30).