

12 The wizards of deception detection

Maureen O'Sullivan and Paul Ekman

Wizard . . . a sage . . . a person of amazing skill or accomplishment

For many years we have studied individual differences among people in their ability to detect deception from demeanour. Most people do rather poorly in making such judgements. In experimental situations in which someone is either lying or telling the truth half the time, people do little better than chance (i.e., they get average accuracy scores of 50 per cent). This has been found not only in our studies (Ekman and O'Sullivan, 1991), which examined high-stake lies, but also in a wide range of other studies (Malone and DePaulo, 2001).¹ By doing an idiographic analysis, we have discovered some individuals who are highly accurate across different types of deception. This report describes our beginning research with a small group of such 'wizards' of deception detection.

In the late 1980s, we started to administer a test of the ability to detect deception, based on our early work with a group of nurses who had lied or been truthful about their feelings as they watched either a gruesome surgical film or a pleasant nature film (Ekman and Friesen, 1974). When we analysed group performance, we found that most groups – police officers, CIA and FBI agents, lawyers, college students, therapists, judges, etc. – did little better than chance. Yet our objective measurements had demonstrated that there were discernible clues that could have been used to distinguish lying from truthfulness accurately (Ekman, Friesen, O'Sullivan, and Scherer, 1980; Ekman, Friesen, and O'Sullivan, 1988).

We (Ekman and O'Sullivan, 1991) did find one group – Secret Service agents – who as a group, scored significantly better than chance, and slightly more than half of them were highly accurate, scoring 70 per cent or better on the test. At that time, we speculated that there were at least two reasons that the Secret Service agents were better at detecting deception than other groups we had tested. The Secret Service has two different jobs. One is to protect public officials, especially in public venues such as giving speeches in front of many people. We speculated that their practice at scanning crowds sensitised them to the non-verbal clues to deception contained in our deception task.

The other job the Secret Service does is to evaluate threats made against public officials. Most of these threats are made by mentally ill individuals. Secret Service agents must decide whether a threat is the precursor to action or merely the outward manifestation of a pathology that is unlikely to lead to behaviour. Unlike many police officers who seek evidence to support their belief in the guilt of the suspect they have arrested, the Secret Service must keep an open mind about the guilt or innocence of the person whose threat they are investigating. Premature closure about the truthfulness or deceptiveness of others is one of the cognitive drawbacks involved in inaccurate lie detection (O'Sullivan, 2003). Another aspect, of course, is the low base rate of occurrence of assassins. Other law-enforcement personnel deal with groups in which the base rate of crime and guilt is much higher. The relative infrequency of the threats the Secret Service deals with may heighten their sensitivity to relevant cues to truthfulness.

The publication of Ekman's book *Telling lies* (1985), and subsequent media publicity, drew the attention of law-enforcement officers who were interested in improving training on interrogation. Among these was J. J. Newberry, an agent in charge of interview training for the department of Alcohol, Tobacco, and Firearms. He scored 100 per cent on the test available at that time on the ability to discern deception and truthfulness. Over the next few years, Ekman and Newberry formed the Institute for Analytic Interviewing which provides training for a wide variety of law-enforcement agencies and which gives us access to many highly specialised and highly motivated populations.

During this period, Frank and Ekman (reported in 1997) were in the process of developing two other deception scenarios (in addition to Ekman and Friesen's scenario of nurses lying or telling the truth about emotion). In one scenario, men lied or told the truth about their opinion; in the other, men lied or told the truth about whether they had stolen \$50. These scenarios appear to have more applicability to law enforcement training than the earlier emotion scenario. In 1999, Ekman, O'Sullivan, and Frank reported their identification of three different professional groups, each of which was significantly better than chance in recognising truthfulness and deception. These groups were: (1) law-enforcement officials from a variety of agencies who had shown special interest in improving their lie-detection abilities and who had been nominated by their agencies as superior interviewers, (2) federal judges, and (3) forensic psychologists. Because two of these groups had volunteered for such training and one of them was willing to lose pay and to travel some distance from home, we speculated that what all three groups had in common was

a high level of interest in the topic of lie detection and unusual motivation to improve their ability in this area.

Other 'wizards' of deception detection were discovered when a very large group (approximately 1,200) of psychotherapists were given the opinion deception scenario. About 10 per cent of the therapists scored 90 per cent or better on the test. Having data on such a large sample substantiated our impression from the earlier studies of smaller groups that this kind of lie detection 'wizardry' is as statistically infrequent as any kind of genius or talent. We believe it is likely that less than 1 or 2 per cent of selected populations will achieve great accuracy in detecting deception from demeanour. Our report focuses on what we have learned so far about the characteristics of these highly talented lie-catchers.

First, let us provide more detail about the three lie-detection accuracy measures we have used in identifying our wizards.

The Lie-detection measures

The Emotion Deception Judgement Task

The task is based on a black and white videotape of ten Euro-American women lying or telling the truth about the feelings they are having at that moment (Ekman and Friesen, 1974; Ekman, Friesen, and O'Sullivan, 1988). In each of the ten items, the woman is shown being interviewed by a female interviewer, who is not visible on screen and who did not know which film the woman she was interviewing was watching. Each woman said she was watching a pleasant nature film, but only about half of them actually were. The others were watching a man with third-degree burns over much of his body intermixed with close-up views of another person undergoing surgery on his arms.

Extensive behavioural measurements of the videotapes identified a variety of behavioural clues in both verbal and non-verbal channels that distinguish lying from truthfulness (Ekman et al., 1980; Ekman, O'Sullivan, Friesen, and Scherer, 1991), establishing that the honest and deceptive interviews can be discriminated from the information available. The women shown on the videotape were students entering the University of California at San Francisco Nursing School who had received a letter from the Dean of the Nursing School inviting them to participate in the research. To motivate them further, the student nurses were told that their ability to control their emotional reactions would be related to their future success as nurses. And, indeed, there was a significant positive correlation between how detectable the lies of each woman were and the

ratings made of her by her clinical supervisor three years later (Ekman, 1985).

The emotion video was taped in 1970 and the women's hairstyles are dated, but participants have not responded negatively to completing this task, other than to comment on the difficulty of determining whether the women are lying or telling the truth. Although accuracy is close to chance for most occupational groups on all three tests, we have found that the emotion task is the most difficult. It yields the smallest proportion of individuals who are highly accurate in their judgements.

The Opinion Deception Judgement Task

This test was developed by Frank and Ekman (1997) from colour videotapes of twenty men who volunteered to be research subjects in a study about human communication. The men in the opinion test were 18 to 28 years of age and included African-, Asian-, and Euro-Americans. Based on the false opinion paradigm suggested by Mehrabian (1971), the men were asked about their opinions before they were told that the experiment involved lying or telling the truth. The opinion that each man felt most strongly about was the one he was then asked to discuss with an interviewer (Ekman). Some men described their opinion truthfully; others lied, claiming to believe the opposite of their true opinion. Truth-tellers who were believed by the interviewer received a \$10 bonus. Liars who were believed received a \$50 bonus. Liars or truth-tellers who were disbelieved received no money and half of them faced an additional punishment. (See Frank and Ekman, 1997, for more details.)

To verify that the men actually did manifest different behaviours when lying or telling the truth, their facial muscle movements were analysed using the Facial Action Coding System (Ekman and Friesen, 1978) which demonstrated that '80% of the participants . . . could be successfully classified as liars or truth tellers on the basis of the presence or absence of fear or disgust' (Frank and Ekman, 1997, p. 1433). The Opinion Deception Judgement Task consists of one-minute segments from each of ten videotaped interviews. Each item shows a different man. The ten interview segments selected were chosen so as to represent equal numbers of pro and con positions on each of the opinions represented. Half of the interview segments are of men who truthfully described their strongly held opinion; half lied about it. (The opinions discussed were either 'Convicted cold-blooded murderers should be executed', or 'Smoking should be banned in all public places'.) The videotapes showed face-and-shoulder close-ups with full audio. The interviewer could be heard but not seen.

As with the emotion videotape, the merit of this detection task is that each of the interviews contains behavioural and/or verbal content clues to either honesty or deception. The interviewees were highly motivated by money rewards and the threat of a noxious punishment. In addition, they were interviewed about opinions they were passionately interested in. Consequently, the subjects displayed emotions consistent or inconsistent with what they were saying, so there were sufficient clues in the videotaped interviews to allow honesty or lying to be detected. (Most previous lie-detection research by other investigators has focused on low-stakes lies (e.g., polite, 'white' lies) that may not cause any emotional arousal in those telling them, thereby providing few, if any, clues for observers to detect (DePaulo, Lindsay, Malone, Muhlenbruck, Charlton, and Cooper, 2003).) Moreover, most investigators have not taken the necessary step of obtaining behavioural measures to document that there are discernible clues that could be used to distinguish lies from truthfulness.

The Crime Deception Judgement Task

This deception judgement task was also developed by Frank and Ekman (1997). Ten men (different from the ten men used in the Opinion Deception Judgement Task) lied or told the truth about whether they had stolen \$50. As in the opinion task, the men received a significant reward if they were believed. If they lied and the interviewer believed their protestations of innocence, they kept the \$50 they had stolen. If they were innocent and the interviewer believed them, they received \$10. If they were disbelieved (whether they were innocent or stole the money), they did not get to keep the \$50, and they were threatened with a noxious punishment. (See Frank and Ekman, 1997, for more details.) As in the Opinion Deception Judgement Task, the men were highly motivated to succeed at the task, whether it was lying or telling the truth. Those who lied showed visible and measurable behavioural signs of cognitive and emotional arousal. Each man was shown being interviewed for about one minute. Half the men were lying and half were telling the truth.

How were the experts identified?

We disseminate our research findings for use by law-enforcement and other professionals through workshops on behavioural cues to deceit and the precautions necessary in interpreting such cues. These workshops vary in length from one hour to five days. The content of the presentation for the longer workshops includes techniques for interviewing developed by J. J. Newberry based on Fisher and Geiselman's (1992) cognitive

interviewing paradigm. The section on detecting deceit is based on Ekman's research on behavioural clues to deception. The one-hour workshops focus on the pitfalls involved in detecting deception. The Opinion Deception Judgement Task is administered at the start of the workshop so participants can gauge their own lie detection ability. We believe this has served to motivate their attention to the training that is offered.

After completing the opinion task, which takes about fifteen minutes, the workshop participants are given the correct answers and asked to report to the group (by a show of hands) how many items they got right. Usually, only one participant in a hundred indicates that he or she got a score of 90 per cent or better. We then ask those who got scores of 90 per cent or better, and who are willing to help in our research, to give us their names and contact information.

Not everyone who is highly accurate has agreed to take further tests as described below. In subsequent reports, we will describe the differential volunteer rate of different occupational groups and the reasons that may explain those differences.

Who are the experts?

Individuals in groups such as those reported in our earlier lie detection accuracy research (Ekman and O'Sullivan, 1991; Ekman et al., 1999) are not included in the Wizards Project because, at that time, we did not identify the participants as individuals. We only gathered group data. The groups in the current research containing highly accurate lie-catchers, who are being studied individually, include therapists, law-enforcement personnel, judges, lawyers, arbitrators, and artists.

The qualification process

Each of the three lie-detection measures described earlier contains ten items, each having a binary choice – lying or truthful. By chance alone, an observer should be able to identify five of the ten items correctly. Based on the binomial distribution, a score of nine out of ten correct, by chance alone, will occur about 1 per cent of the time. A score of eight out of ten will occur, by chance alone, 5 per cent of the time.

We decided that initial qualification for the Wizards Project would require achieving the very high score of 90 per cent on the Opinion Deception Judgment Task. Having achieved that cut-off score, the crime and emotion videotapes were then mailed to the participants. At first we set the criterion very high. If the participants achieved scores of 80 per cent on *both* the crime and the emotion tests, they would be classified as

'expert' lie detectors or 'wizards'. An additional consent form was then sent to them in which they agreed to be interviewed about their answers to the tests, to talk with us about aspects of their life and work, and to complete a battery of psychological tests.

What is the probability that wizards selected by the criteria described above could have achieved their scores by chance alone? Essentially non-existent. On the basis of chance alone, the combinatorial probability that any single individual would achieve a score of 90 per cent on the first lie-detection test and scores of 80 per cent or better on two other tests is twenty-five in a million. So, it is highly unlikely that we have any false positives, that any of our wizard lie-catchers are incorrectly classified as highly accurate. It is quite likely, however, that we have many false negatives, that other expert lie detectors were falsely categorised as non-experts. The false negatives may be attributed to the fact that we do not provide our participants with many of the opportunities that are normally available in an interview – a longer sample of behaviour, the ability to formulate the questions and follow-ups, the opportunity to use role-playing and other behavioural stratagems, etc. Note, however, that the experts we have identified are able to detect deception, even with these restrictions.

Professional experience and accuracy in detecting deception

At first, we used the very stringent criteria described above for identifying our experts. One in which for all three tests (90 per cent on the opinion test and at least 80 per cent on both the crime and emotion tests) the odds of being classified as an expert by chance alone were extraordinarily low. Using these criteria, we identified fourteen 'ultimate' wizards.

We first started identifying potential experts using our large group of therapists. We found that some of them failed to qualify as experts because, although they scored 90 per cent on the opinion tests, and 80 per cent or above on the emotion test, they failed to achieve a score of 80 per cent or better on the crime tape.

As the study progressed, and we received data from other professional groups, we noticed a complementary pattern among law-enforcement personnel (police, judges, and lawyers) who scored 80 per cent or higher on the crime tape but failed to reach criterion on the emotion tape. Fifteen participants (in addition to the fourteen 'ultimate' experts) achieved scores of 80 per cent or better on two of the deception scenarios, but not on the third. Table 12.1 shows the frequencies with which these participants, classified as therapist or law enforcement² failed to achieve a similarly high score on either the crime or the emotion test.

Table 12.1 *Deception detection inaccuracy of two professional groups on two tasks*

Profession	Deception detection task	
	Crime	Emotion
Therapists	8	1
Law professionals	0	6

Notes: Values are the number of professionals in each group who obtained scores of 70% or less on each test, having scored 80% or higher on the other test, i.e., 6 law professionals achieved scores of 80% or higher on the crime task, but all of them received 70% or less on the emotion task.

The chi-square ($\chi^2 = 11.429$ $p < .00072$; with Yates' correction $\chi^2 = 8.136$ $p < .0043$) is highly statistically significant. It appears that the type of lie one least often encounters is the one that is responsible for failure to qualify – therapists on crime, law enforcement on emotions.

Therapists, while keenly observing their clients for areas of emotional and personal significance would not ordinarily have reason to be concerned about whether their clients were criminal, nor would they have experience with the kinds of stratagems people might use to avoid having their crimes detected. Conversely, law-enforcement personnel are in the business of detecting lies about crimes and would have more familiarity with that kind of deception. Lies about feelings would be less important in their day-to-day professional lives. (Recall, however, that our fourteen original 'ultimate' experts were able to identify all three kinds of lies accurately and that the ultimate experts include therapists and law-enforcement personnel.) This finding of seeming professional relevance for the lies correctly identified encouraged us to study this second group of experts as well, to try to understand what differentiates their expertise. By October, 2003, we had identified fourteen 'ultimate' experts, who had obtained 80 per cent or better on all three deception detection tasks, and fifteen 'penultimate' experts who had scored 90 per cent or better on the opinion test and 80 per cent or better on either the crime or the emotion test. We continue to assess potential wizards.

Although we believe that the pattern of differential accuracy shown in Table 12.1 reflects differences in professional experience with different kinds of lies, other hypotheses need to be ruled out. One alternative explanation is that it is not the type of lie that is important, but the

Table 12.2 *Deception detection inaccuracy by sex of target and sex of expert*

Sex of expert	Sex of liar or truth-teller	
	Males interviewed about stealing money (Crime video)	Females interviewed about film viewing (Emotion video)
Female	2	2
Male	5	6

Notes: Values are the numbers of male and female 'penultimate' experts who scored 70% or lower on either the crime or the emotion video. Penultimate experts obtained 90% on the Opinion Deception Detection Task and 80% on either the emotion or crime task. Table 12.2 shows the distribution, by sex, of the penultimate experts on the video on which they did **not** achieve criterion. As can be seen, there is no relationship between the sex of the target and the sex of the expert.

gender of the liar and the lie-catcher. Most of the therapists and all of the people lying about their feelings are female; most of the law-enforcement personnel and all of those lying about stealing money are male. This explanation is less satisfactory, however, since all the penultimate experts, male and female, were highly accurate in detecting lies about opinions told by males (the initial screening task). Further, a chi-square partitioning the penultimate expert data in terms of the gender of the liars and truth-tellers and the gender of the lie-catchers ($\chi^2 = 0.024$ $p < .877$) was not significant. These data are given in Table 12.2.

The protocol

After the experts are identified, they meet with one or both of the experimenters (O'Sullivan and Ekman) to review their responses to each of the three videos. They are instructed to say aloud anything that occurs to them as they watch the video. They are encouraged, in other words, to 'think aloud' (Ericsson and Simon, 1998). This procedure was usually the first activity we did with the experts since we did not want to influence their recall or their reporting of the process they used in detecting deception. We said very little during this initial phase, interrupting only to encourage the expert to 'talk aloud' if they were merely watching the tape, or asking clarifying questions when remarks were made which we did not understand. After reviewing all three tapes, we sometimes went back over the tapes and called the experts' attention to aspects of the tape

that they had not mentioned, aspects of the interviews that we knew contained important information about the truthfulness of what was being said, and asked them for their comments on that part of the videotaped interview. This initial review procedure lasted from one to two hours and was tape-recorded and transcribed, a lengthy and difficult procedure since there are four or five voice sources.³

Ordinarily, the next step is to conduct a semi-standardised interview about personal and life history information. Early on, we found that this interview was adequate for gaining factual biographical information and, occasionally, information about career choices, mentors, and the like. But, often, the experts were overly humble about their skills or provided only biographical facts. This information was also tape-recorded and transcribed.

The experts who were interviewed first were law enforcement personnel with whom we were conducting week-long training courses. During this time, we would eat and drink together and the police officers would talk about each other, tell 'tall tales' about their police exploits, and be far more revealing of themselves and their colleagues than they had been in the formal semi-structured interview. Because of the richness of the data gathered in this way, we are trying to obtain more naturalistic observations for all of our experts, by visiting them at home, or at their job, and by talking with colleagues and family members who might be more forthcoming. In the course of these social interactions with the experts, we make detailed notes as soon after our encounters as possible.

In addition to the think-aloud procedure and the gathering of interview and observational data about their personal and professional lives, a series of standardised psychological tests will be administered to the experts. These will include a personality inventory such as the NEO, a short measure of verbal comprehension as an indirect measure of IQ, a measure of attributional or cognitive complexity, and perhaps measures of cognitive style and life themes.

Limitations of the identification protocol and this report

We are aware that our procedure will select for a particular kind of lie-catcher. Our task limits the lie detector to watching a one-minute videotape of a liar or truth-teller being interviewed by someone else. The sample of behaviour is quite limited (one minute), the observer does not have the opportunity to obtain a baseline sample of behaviour that is known to be honest to use as a comparison, the observer is limited to merely watching the behaviour on the videotape. Highly skilled interviewers who have learned over the years how to use their appearance and personality to best effect, or how to interview different personality types

with different techniques, or who have developed particularly effective interview strategies for use in detecting deception may be handicapped by the protocol we use which does not allow them to use these talents and techniques. On the other hand, in earlier research (Ekman et al., 1999) interviewers identified by their agencies as superior interviewers also were identified as superior lie detectors by one of the measures used in the present study. Although not all kinds of expert lie detectors will be detected by our methods, many of them will be.

We have only started interviewing the wizards. As of July, 2003, only half of the ultimate experts and none of the penultimate experts have been interviewed. We can offer no definitive findings (other than the relationship between professional experience and differential expertise described above), but a description of the demographic characteristics of the experts, beginning evidence for some of our hypotheses and some overall impressions of these wizards may be of interest.

Demographic characteristics of the ultimate and penultimate experts

Sex

Although we did not collect precise data on the gender composition of our samples, in most of our samples the number of males was notably greater than the number of females. Nonetheless, 10 of the 29 experts are female, which is a larger proportion of women than would be expected by chance alone. This finding is surprising since neither we (Ekman and O'Sullivan, 1991) nor others (Anderson, DePaulo, and Ansfield, 2002) have found a sex difference in lie-detection accuracy. On the other hand, women often perform better than men on non-verbal accuracy tasks (Hall, 1990) and one kind of non-verbal ability, recognising micro momentary facial expressions, is correlated with lie detection accuracy (Ekman and O'Sullivan, 1991; Frank and Ekman, 1997). The seeming over-representation of women in our expert group may be the result of women being more likely to volunteer to participate in a research project and to be more conscientious in completing their agreement to participate. The issue of sex differences in expert levels of accurate lie detection is one of the many questions we hope to address in future analyses of this data set.

Age

Consistent with the data described earlier suggesting that experience is important in recognising a wide variety of lies, most of our accurate lie-catchers are middle aged. The youngest is a female law student in her

twenties; all the others are between 40 and 60. Given the depth of information and understanding of the personalities and behaviours of our widely varying truth-tellers and liars, this age range is understandable. Our young expert may provide a unique glimpse of ways in which a sophisticated understanding of others develops early in life.

Hypotheses suggested by previous research

Non-verbal behavior

In previous work (Ekman and O'Sullivan, 1991; Frank and Ekman, 1997) we found a significant correlation between accuracy in lie detection and accuracy in interpreting micro momentary facial expressions of emotion. We also found (Ekman and O'Sullivan, 1991; O'Sullivan, 2000) that more accurate lie detectors report using non-verbal as well as both verbal and non-verbal cues more frequently than less accurate lie detectors.

Based on these findings, we predicted that the experts who were accurate on the deception detection tasks would use non-verbal cues in making their assessments of truthfulness. In all of the think-aloud interviews we have conducted so far, the experts spontaneously refer to the non-verbal behaviour of the people they are watching 'He looks like a deer in the headlight of a car' (and the expert showed on his face the open, transfixed look of a deer). 'At that point, he smirked, and it didn't make sense that he would do that.' Every expert interviewed so far has described the non-verbal behaviour of the liar or the truth-teller, as well as the consistency or inconsistency between verbal and non-verbal behaviour on some, if not all, of the individuals they are assessing.

One of the things we hope to achieve with this study is a description of the non-verbal paradigms that expert lie detectors use in understanding others. Do they have a single strategy that they tailor for each person? Or do they use different strategies for different kinds of lies? Are there some non-verbal (and verbal cues) that experts use more than non-experts, or use in a different way, or in different combinations?

Verbal behaviour

For many years, many researchers (O'Sullivan, Ekman, Friesen, and Scherer, 1985; Köhnken, 1987; Sporer, 1997; Yuille, Marxman, and Cooper, 1999; Ekman, 2001; DePaulo et al., 2003; Newman, Pennebaker, Berry, and Richards, 2003) have argued that the verbal content of speech is important in differentiating deception and truthfulness. This occurs because the cognitive difficulty of constructing an alternative

reality is revealed in reduced word complexity, less immediacy, more distancing, as well as vague and general language. In addition to their non-verbal facility many of the expert lie-catchers show an exquisite sensitivity to the nuances of language use.

Childhood experiences

Another consistency among the handful of experts who have been interviewed so far is that most of them have had unusual childhoods. Some of them did not speak English until grade school, some are children of alcoholics, some had working mothers, when not many children their age did. Most of them report being aware of changes in the emotional levels of those around them from a very early age.

Motivation

In our early studies of groups of experts (Ekman and O'Sullivan, 1991; Ekman et al., 1999), we speculated that superior lie-catchers were distinguished by their motivation to do well at the task as evidenced by their taking a workshop on the topic of detecting deception when others in their professions did not. All the experts in the Wizard Project are like the well-motivated lie detectors in our earlier studies. They, however, are even more highly motivated or, perhaps, more responsible and/or conscientious. Most attendees at our workshops are very enthusiastic about participating in the project at the beginning. But time and distance dampens their ardour. Although some groups had high response rates, in all the professional groups, many well-meaning, but busy, professionals failed to respond to our request for their help, although they had initially agreed to do so. The ultimate and penultimate experts did respond. We hypothesise that they will be more conscientious than a control group on a standardised personality measure.

Habitual practice

Ericsson (1996) in his review of expertise in professions as widely varying as grand master chess champions and concert violinists argued that the one characteristic that distinguishes the highly expert from the merely good was an extended period of intensive practice, usually lasting ten years, at the start of their careers. Recently, Brassington (2003) reported similar findings with prima ballerinas and members of the corps de ballet. Prima ballerinas reported more practice, visualising more about their

performance, more planning. We will be assessing these planful, intellectual habits in our experts as well. Our observation, at this point, having interviewed only a third of them, is that an intense focus and investment in their performance, and concentrated attempts to improve it, are characteristic of most, if not all, of the wizards.

Other hypotheses

As noted above, our protocol has limitations. Those limitations may have skewed our sample. Our experts certainly do not represent all the possible kinds of expert lie detectors there are. Champion poker players, for example, might be different in personality characteristics as well as in their lie-detection strategies. Nevertheless, even with such limitations, the interviews conducted so far provide some interesting preliminary observations. These include the following:

- Many of our experts seem to be introverted and observant, rather than extroverted and active. Although some of our experts are extraordinarily, even excessively outgoing, most are not. Most are quiet, reserved, and observant. This hypothesis will be evaluated with a standardised personality inventory.
- When necessary, a surprising number of our experts report acting as a situation demands, whether that means being low-key and inoffensive or dominant and take-charge. Their ability to understand others seems to coexist with an ability to act the role or the behaviour that is required in a given situation. The understated, low-key, observant lie detectors ride Harleys, wear ten-galloon hats and high-heeled cowboy boots, or dress the part of the sailor-boy or the army captain when necessary. We hypothesise that the expert lie detectors will relate examples of noteworthy role-playing in their professional and personal lives significantly more often than the members of the control group will.
- Although a few of our experts do not have college degrees, most of them do and some have postgraduate degrees. The group, as a group, is surely above the mean in intelligence, but this will be assessed through objective measures.
- The experts interviewed so far seem to observe and describe people in a more complex fashion and with a thoroughness that others do not. Not only do they seem to have their own personality theories, but they also show a nuanced understanding of social, racial, and individual differences. Although some of the experts expressed many of the same stereotypic beliefs of non-experts, the wizards seem able to put these beliefs aside as they attempt to understand this particular young white woman, this particular black man.

- Most of the experts have had a familiarity with a wide array of very different types of people. One Los Angeles County Sheriff worked with psychiatric patients early in his career, as well as with prostitutes of many sexual persuasions, including transsexuals, but he goes to garage sales with his daughter and helps her find Daisy Duck collectibles. A labour attorney has dealt over the course of his career with people having the lowest level of education and intellectual capacity to the highest, from ore boat captains or wildcat coal-mine operators in the Appalachians to CEOs of multinational corporations.
- So far, our expert lie detectors show no discernible pattern with respect to political and religious beliefs. At this point, an equal number of them are Republicans and Democrats. Similarly, although many are extremely religious or very active in their church, some are agnostic, atheistic, or non-practising.
- How do the experts construe the personalities of the people they observe? What dimensions characterise their observation? How important is their assessment of the kind of person they are observing as a base-line for evaluating the non-verbal cues they are so sensitive to? How do they know when to attend to a non-verbal cue and when to ignore it? These are questions we are exploring.
- Some of the wizard lie-catchers have hobbies or vocations that either honed their acute sensitivity to non-verbal cues or became hobbies because of their inherent talent in that area. One expert is a bird watcher, another is an artist who does detailed paintings of facial expressions of emotion, another is a hunter known for his tracking ability and his persistence. We do not yet have the control group information to know, however, whether these observations are relevant or serendipitous. Which brings us to our final consideration: What is the control group?

The control group

One of the dilemmas of this project has been the problem of defining an adequate control group. Given the age, professional accomplishments, and talents of our wizard lie-catchers, the task of finding an adequate control group was daunting. The members of the control group would have to be non-expert lie detectors, by our criterion, but also be interested to participate in the research project. They would have to be similar to the experts in terms of social class, educational level, geographic location, and age.

There was no pre-existing group that met these requirements, so we decided to use each expert's spouse or close family member as his or her

control. In most cases, the family member is interested in the project, and in their spouse's involvement. In some cases, before we had decided to use spouses as controls, husband or wives of the potential expert took the test too, out of personal interest. In one case, the wife was the target and did not qualify. Her husband took the test and did qualify!

The reasons recommending this strategy of using the spouses as the control group are that the spouse would usually be similar in age, as well as social and educational level. And since they are still married, they would both be living in the same part of the country. Also, since most of the experts are at least middle-aged, they would have been married for some time. There is anecdotal speculation, as well as research evidence, that married people become more similar over time, at least in terms of emotional convergence (Anderson, Keltner, and John, 2003). This similarity provides a strong comparison group, so that any differences we find are likely to be meaningful ones. We will, however, be exploring the utility of other control groups as well.

Conclusion

For Olympic athletes, talent is a necessary, but not a sufficient, condition for athletic excellence. The development of even the greatest gifts takes practice, feedback, and motivation. We speculate that the same is true for the development of the ability to understand others. Almost all the experts we have interviewed have exhibited high levels of motivation to improve their ability to understand others and to seek feedback about their performance. These characteristics are integrated into the everyday professional lives and personalities of the experts we have interviewed so far. They are not 'tricks' that can be taught in a four-hour seminar. Hopefully, however, our study of the wizards of deception detection will enrich our understanding of how human beings communicate and understand truthful and deceptive behaviour and provide information to those of us seeking to detect deception more accurately. We have suggested to government officials that these wizards are a resource that might be consulted on cases of extraordinary importance when there are few other means of evaluating truthfulness in a limited time frame.

REFERENCES

- Anderson, C., Keltner, D., and John O. P. (2003). Emotional convergence between people over time. *Journal of Personality and Social Psychology*, 84 (5), 1054-68.
- Anderson, D. E., DePaulo, B. M., and Ansfield, M. E. (2002). The development of deception detection skill: A longitudinal study of same-sex friends. *Personality and Social Psychology Bulletin*, 28 (4), 536-45.

- Brassington, G. (2003). Mental skills distinguish elite soloist ballet dancers from corps dancers. Paper presented at the annual meeting of the Western Psychological Association Convention, Vancouver, B. C., May, 2003.
- DePaulo, 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. (1985). *Telling lies: Clues to deceit in the marketplace, politics, and marriage*. New York: Norton.
- (2001). *Telling lies: Clues to deceit in the marketplace, politics, and marriage* (3rd edn). New York: Norton.
- Ekman, P., and Friesen, W. V. (1974). Detecting deception from body or face. *Journal of Personality and Social Psychology*, 29, 288-98.
- Ekman, P., and Friesen, W. V. (1978). *Facial action coding system*. Palo Alto, CA: Consulting Psychologists Press.
- Ekman, P., Friesen, W. V., O'Sullivan, M., and Scherer, K. R. (1980). Relative importance of face, body, and speech in judgments of personality and affect. *Journal of Personality and Social Psychology*, 38 (2), 270-7.
- Ekman, P., Friesen, W. V., and O'Sullivan, M. (1988). Smiles when lying. *Journal of Personality and Social Psychology*, 54, 414-20.
- Ekman, P., and O'Sullivan, M. (1991). Who can catch a liar? *American Psychologist*, 46 (9), 913-20.
- Ekman, P., O'Sullivan, M., and Frank, M. G. (1999). A few can catch a liar. *Psychological Science*, 10 (3), 263-6.
- Ekman, P., O'Sullivan, M., Friesen, W. V., and Scherer, K. R. (1991). Invited article: Face, voice, and body in detecting deceit. *Journal of Nonverbal Behavior*, 15 (2), 125-35.
- Ericsson, K. A. (1996). The acquisition of expert performance: An introduction to some of the issues. In K. A. Ericsson (ed.), *The road to excellence: The acquisition of expert performance in the arts and sciences, sports, and games* (pp. 1-50). Hillsdale, NJ: Lawrence Erlbaum.
- Ericsson, K. A., and Simon, H. A. (1998). How to study thinking in everyday life: contrasting think-aloud protocols with descriptions and explanations of thinking. *Mind, Culture and Activity*, 5 (3), 178-86.
- Fisher, R. P., and Geiselman, R. E. (1992). *Memory enhancing techniques for investigative interviewing: The cognitive interview*. Springfield, IL: Charles C. Thomas.
- Frank, M. G., and Ekman, P. (1997). The ability to detect deceit generalizes across different types of high-stake lies. *Journal of Personality and Social Psychology*, 72 (6), 1429-39.
- Hall, J. A. (1990). *Nonverbal sex differences: Accuracy of communication and expressiveness*. Baltimore, MD: Johns Hopkins University Press.
- Köhnken, G. (1987). Training police officers to detect deceptive eyewitness statements: Does it work? *Social Behavior*, 1, 1-17.
- Malone, B. E., and DePaulo, B. M. (2001). Measuring sensitivity to deception. In J. A. Hall and F. J. Bernieri (eds.), *Interpersonal sensitivity: theory and measurement* (pp. 103-24). Mahwah, NJ: Lawrence Erlbaum.
- Mehrabian, A. (1971). Nonverbal betrayal of feeling. *Journal of Experimental Research in Personality*, 5 (1), 64-73.

- 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), 665–75.
- O'Sullivan, M. (2000). Why I think you're lying. Paper presented at the International Congress of Psychology. Stockholm, Sweden, July 2000.
- (2003). The fundamental attribution error in detecting deception: The boy-who-cried wolf effect. *Personality and Social Psychology Bulletin*, 29, 1316–27.
- (in press). Detecting deception and nonverbal communication: Why most people can't 'read' others, but a few can. In R. Riggio, and R. Feldman (eds.), *Applications of Nonverbal Communication*. Mahwah, NJ. Erlbaum.
- O'Sullivan, M., Ekman, P., Friesen, W., and Scherer, K. R. (1985). What you say and how you say it: The contribution of speech content and voice quality to judgments of others. *Journal of Personality and Social Psychology*, 48 (1), 54–62.
- Sporer, S. L. (1997). The less traveled road: Verbal cues in deception detection in account of fabrication and self experiences events. *Applied Cognitive Psychology*, 11 (5), 373–97.
- Yuille, J. C., Marxman, D., and Cooper, B. (1999). Training investigative interviewing: Adherence to the spirit as well as the letter. *International Journal of Law and Psychiatry. Special issue: Current Issues in Law and Psychiatry*, 22 (3–4), 323–36.

NOTES

We wish to thank Leora Lubliner for her extraordinary contributions to the administrative organisation and data analysis involved in the initial phase of the Wizard Project. We also offer our sincere gratitude to the wizards who have been so generous with their time and talent.

1. We (Ekman, 1985; 2001; O'Sullivan, 2003; O'Sullivan, in press) have described numerous reasons for most people's inability to detect deception.
2. One law professor, who had a master's in counselling and taught an interviewing/self-knowledge class in a law school, was classified as a therapist rather than as law enforcement because the bulk of his work was psychological rather than legal.
3. The voice of the expert lie-catcher, the interviewer or interviewers, and, in the background, the voice of the liar/truth-teller and the person interviewing him or her.