Finding a needle in a haystack: toward a psychologically informed method for aviation security screening

Article (Accepted Version)

Running head: Detecting deception during security screening

Finding a needle in a haystack: Towards a psychologically-informed method for aviation security screening

Thomas C. Ormerod[1]*

Coral J.Dando[2]

[1]Department of Psychology, University of Surrey, UK.

[2]Department of Psychology, University of Wolverhampton, UK.

*Correspondence to: t.ormerod@surrey.ac.uk

Published in Journal of Experimental Psychology: General, 2014.

Abstract

Current aviation security systems identify behavioural indicators of deception to assess risks to flights, but they lack a strong psychological basis or empirical validation. We present a new method that tests the veracity of passenger accounts. In an *in-vivo* double-blind randomised-control trial conducted in international airports, security agents detected 66% of deceptive passengers using the veracity test method compared with less than 5% using behavioural indicator recognition. As well as revealing advantages of veracity testing over behavioural indicator identification, the study provides the highest levels to date of deception detection in a realistic setting where the known base rate of deceptive individuals is low.

"You have to measure whether what we're doing is the only way to assure ... safety. And you also have to think are there ways ... that are less intrusive" (Barak Obama, Lisbon, November 2010).

Identifying threats presents a huge challenge to those tasked with ensuring public safety, and to psychologists developing methods for detecting deception. However, the news from both arenas is not good. Since the events of September 11 2001, billions of dollars have been invested in aviation security procedures designed to detect threats to airplanes (US-GAO, 2011), but the effectiveness of these procedures has been questioned (Weinberger, 2010). More recent events such as the 2009 attempted bombing of Flight NW253 to Detroit suggest we still lack effective ways of identifying threats to public safety. Threat detection procedures typically involve looking for individuals who display behaviours thought to be indicators of deception, particularly behaviours shown by the perpetrators of previous attacks. The effectiveness of behavioural indicator approaches has never been tested in a large-scale field trial. However, a meta-analysis of laboratory studies that used behavioural indicators to discriminate deceivers from truth-tellers revealed a mean rate for correct identification of 54%, only marginally above chance (Bond & de Paulo, 2006).

In this paper, we present a new procedure for aviation security screening that is based, not on behavioural indicators selected from previous incidents, but on testing the veracity of passengers' verbal accounts. Our approach takes techniques derived from psychological theory and shown in recent laboratory studies to yield promising rates of deception detection, and integrates them into a comprehensive procedure for detecting threat. We then compare the effectiveness of behavioural-indicator and veracity-testing approaches in an in-vivo empirical evaluation conducted with

passengers departing on flights at international airports during routine security screening.

Approaches to Detecting Deception

The majority of published research on detecting deception has set out to identify indicators in human behaviour that can discriminate deceivers from truth-tellers. Behavioural indicators of deception fall into two main categories: physical behaviours relating to demeanour (e.g., nervousness; aggression) and/or actions (e.g., eye contact, fidgeting); and verbal behaviours relating to the nature and production of speech (e.g., hesitations, use of pronouns). As noted above, behavioural indicator approaches typically yield low discrimination rates (Bond & de Paulo, 2006). However, there is some evidence that counts of illustrators (i.e., hand movements to indicate content or prosody) can provide reasonable levels of discrimination (de Paulo et al, 2003), but the range of practical contexts in which illustrators can be used is limited (e.g., real-time detection of differences in the use of illustrators is likely to prove impractical). According to Levine (2010), the slight but above significance rate of 54% successful detection arises, not because judges of deceptive behaviour have some degree of competence at identifying relevant behavioural indicators, but because there are generally a few deceivers in any study who are particularly poor at masking their lies.

Low rates of deception detection from behavioural indicators arise, according to Levine, Kim and Blair (2010), for four reasons: a lack of indictors with predictive validity; naive beliefs in the predictive validity of certain indicators (e.g., avoidance of eye contact); ignoring information that may indicate deception (e.g., failing to spot inconsistencies in an account); and truth bias, that is, a predisposition to assume the truth of another person's account. In their 'Dangerous Decisions Theory', Porter,

Gustaw and ten Brinke (2010) argue that an initial schema formed from misinterpreted behavioural indicators biases people's judgements about deception, leading to irrational decision-making in the face of contradictory evidence. Despite these concerns, it has been suggested that deception research has "been characterized by a myopic focus on the internal psychological states and corresponding non-verbal behaviors of liars and has failed to adequately consider the situation and context in which truths and lies are told" (Blair, Levine, & Shaw, 2010, p. 423). As we argue below, the same focus is apparent in current security screening practices.

Veracity testing offers an alternative approach to detecting deception that focuses, not on displayed behavioural characteristics of deceivers, but on the nature of the verbal exchange between the sender (the individual attempting to deceive) and the receiver (the individual attempting to detect deception). Recent laboratory studies have revealed five aspects of verbal exchanges that discriminate deceivers from truth-tellers. First, some of the most successful deception detection methods use evidence to challenge accounts during interviews. Evidence-based methods yield up to 75% detection with scripted questions (Blair, Levine, & Shaw, 2010; Levine, Shaw & Shulman, 2010) and 68% with questions created in real time (Dando & Bull, 2011; Dando, Bull, Ormerod, & Sandham, 2013).

Second, questioning styles that elicit rich verbal accounts are also effective in discriminating between truth-tellers and liars (Milne & Bull, 1999; Oxburgh & Dando, 2011; Oxburgh, Myklebust, & Grant, 2010). Open questions do not constrain responses, but necessitate the provision of expansive answers. Importantly, answers to open questions commit passengers to an account of the truth concerning issues such as identity, background, and previous, current or future activities.

Third, tests of expected knowledge, which compare the content of what someone says with information already known, are useful for detecting deception (Blair, Levine, & Shaw, 2010). For example, if you claim to have studied at Oxford University, it would be reasonable to expect you to know how to travel on public transport from the train station to your college. Lack of knowledge and an inability to explain its absence, or a marked change in verbal behaviour when providing answers, may suggest that the information supplied initially may not be veridical.

Fourth, interviewing methods that restrict the verbal manoeuvring of deceivers are also shown to be effective (e.g., Dando & Bull, 2011; Taylor et al., 2013). Verbal manoeuvring involves the strategic manipulation by deceivers of verbal content and delivery, which is intended to control a conversation in order to avoid detection. The quantity of verbalisations produced by deceptive individuals (measured in terms of number of words), and the information content of their verbalisations, tend to vary according to the nature of a verbal exchange. Specifically, deceptive individuals tend to be as verbose as truthful individuals when they are in control of the conversation (e.g., during early exchanges), and they tend to produce as much unsolicited information (and sometimes more) than truth-tellers. However, deceivers become less verbose and deliver less information than truth-tellers when their accounts are being challenged under questioning (Dando et al., 2012).

Fifth, procedures that raise the cognitive load faced by an interviewee typically yield better rates of discrimination between deceivers and truth-tellers (Walczyk, Igou, Dixon & Tcholakian, 2013). For example, asking unanticipated questions during interviews has been shown to raise the cognitive load of deceivers more than truth-tellers, leading to higher detection rates (Vrij et al, 2009).

The techniques described above provide building blocks for constructing an effective method of detecting deception during security interviews. Psychological research has been used to good effect to design practical methods for interviewing witnesses and victims, notably the cognitive interview (see Fisher & Geiselman, 1992). In our view, the same approach can be adopted to designing methods for interviewing to detect deception, but to date no composite psychologically-based methods have been designed for interviewing those suspected of wrongdoing (suspects) outside of laboratory studies. Of course, aviation security interviews are not suspect interviews, but the same requirement to discriminate between truth-tellers and deceivers arises.

Aviation Security Screening to Detect Deception

Most current aviation security procedures rely on the identification of behavioural indicators (e.g., Reddick, 2004; British Security Industry Association, 2008). A common method for screening airline passengers prior to embarking on long-haul flights involves the detection of 'suspicious signs' during a short interview between security agent and passenger (Martonosi & Barnett, 2006). In the interview the agent asks a series of scripted security-related questions that are the same for every passenger. During questioning, agents look for indicators, which are typically behaviours associated with previous security incidents. These signs focus on aspects of a passenger's verbal and non-verbal behaviours, disposition (e.g., nervousness) and appearance (e.g., inappropriate dress for the intended trip) that may be indicators of deceit or threat[1].

In the light of the psychological literature that reveals problems with behavioural indicator approaches, we have developed a new security screening

[1] In the interests of national security we cannot provide in full the specific behavioural indicators that are used in the suspicious signs method.

method, which we call Controlled Cognitive Engagement (CCE). The name refers to the decision-making skills used by the security agent to control an interview so that a passenger provides information that can be tested for veracity. CCE embodies each of the five techniques shown in laboratory studies to improve deception detection rates: Use of evidence; tests of expected knowledge; effective questioning styles; observation of verbal manoeuvring; and asymmetric cognitive loading. Information revealed in the responses to open questions by the passenger in a CCE security interview is used by the agent to construct questions that provide tests of expected knowledge. The interview procedure has phases, but the questions within each phase are not scripted. Instead, agents are trained to use a question construction algorithm that allows the generation of in excess of 1.8m different questions. Therefore, the precise questioning cannot be anticipated by the passenger. CCE is designed to raise the cognitive load faced by deceptive passengers (based on the assumption that passengers who pose a threat to a flight will have to be deceptive if questioned appropriately). At the same time, cognitive load is minimised for legitimate passengers, who experience what appears to be a friendly and informal conversation, albeit that the conversation is managed. That is, the security agent is seamlessly directing the course and progress of the conversation to fulfil specific psychological and practical objectives (e.g., Shepherd, 2007).

CCE is applied during a short interview between security agent and passenger, as follows: An initial phase of rapport building establishes a baseline for the passenger's verbal behaviour, in which they respond to neutral questions that anyone, regardless of intent, could answer truthfully. Passengers are then asked open

questions. Agents develop questions in real time using a proprietary method[2] that combines one or more selected topic dimensions (e.g., education, family, travel, work, skills, interests) and temporalities (past, current, future), and varying question styles, purpose and length, to derive questions that cannot be predicted in advance. The account given by a passenger in response to an open question is then tested for veracity, using focused 'test' questions that seek information the passenger should possess if their account is true. Agents assess the veracity of accounts from the answers given by passengers to questions. Although agents can quickly access Internet resources using a hand held device to check the accuracy of some answers given by passengers, they are discouraged from doing so. Instead, they are trained to look for changes in the depth and quality of answers during the interview (i.e., as a deceptive passenger becomes aware that their account is being assessed or challenged). Cycles of open and test questions repeat with different topics and temporal domains, after which the agent makes a decision regarding the passenger's risk status.

Evaluating the Effectiveness of Aviation Security Screening Procedures

Recent research highlights the importance of context in assessing the effectiveness of methods for detecting mal-intent (Blair, Levine, & Shaw, 2010). Human behaviour is inherently determined by the situation in which actions arise (Ross & Nisbett, 1991). To date, there have been no large-scale field trials of aviation security screening methods. Some proponents of behavioural indicators (e.g., Ekman, 2009) have noted the difficulty of trialling approaches in controlled studies that cannot involve genuine threat (Weinberger, 2010), because of the difficulty in a mock passenger study of creating the high stakes faced by perpetrators of real attacks. Yet,

[2] As with the suspicious signs of current screening methods, the precise algorithm underlying the proprietary question design method of CCE cannot be published for reasons of national security.

as Blair, Levine, and Shaw point out, the effectiveness of any method for detecting deception is likely to be influenced strongly by the context in which it is applied.

To address the gap in knowledge about how security screening methods fare in real task environments, we conducted the first *in-vivo* evaluation of a suspicious signs method conducted in a real airport, and compared it to CCE. The study reported below provides a randomised-control, double-blind field trial of these aviation security-screening methods. We collected data at five international airports during routine security activities, in which mock and real passengers passed through a security interview prior to check-in for long-haul flights. Mock passengers were incentivised to attempt to pass through security undetected while giving untruthful answers during the screening interview. We hypothesised that security agents using CCE would detect more mock passengers than agents using the suspicious signs method. We also hypothesised that interviews using the methods would be distinguished by the verbal behaviours of both passengers and agents: in CCE interviews, agents would speak less and passengers more, compared with suspicious signs interviews; CCE interviews would yield more information from passengers than suspicious signs interviews; the methods would differ in the question types used by security agents. Finally, we hypothesised that deceptive passengers interviewed under the CCE method would show a reduction of verbal content (number of words and information items) in their answers as the interview proceeded.

Method

Participants

Security agents. Ninety-six males and 64 females (M_{age} = 37.4, *SD* = 12.73) from a workforce of 866 staff participated as part of normal working but were free to

withdraw from study participation at any time. All staff had received training in suspicious signs screening (two weeks classroom instruction and one week on-the-job training), and had an average of 4.9 years' experience (*SD* = 2.81) with the method. Agents were aware that tests of screening effectiveness would be conducted during the eight-month trial, but were blind to the presence of mock passengers. Written consent was obtained from agents to audio-record interviews. Staff were randomly allocated to suspicious signs (83 agents) and CCE (77 agents) methods. Staff allocated to suspicious signs received no further training. Staff allocated to CCE underwent CCE training, comprising one week of classroom instruction (lectures, role-play exercises, and continuous assessment), and one week on-the-job training.

Passengers. Two groups participated (113 male & 91 female in each group): mock passengers (mean age 45.7 years, SD = 5.14) and genuine passengers (mean age 43.5 years, SD = 4.77). Nationalities matched across groups were: American (43), British (116), German (16), Irish (5), Swiss (11), and Canadian (13). Mock passengers were recruited by advertisement. Each was paid £60 to participate, with an additional £60 if they passed through the security interview undetected as deceptive. During days when audio-recordings were taken, researcher approached each passenger after their security interview, seeking permission for their interview to be included in the research. The sample of genuine passengers was selected from 7000 security interviews audio-recorded during the evaluation, matched to mock passengers by age, gender, nationality and first language.

Procedure

One week before participation, mock passengers provided full names, gender; date and place of birth, passport number, and home and work addresses. Each participant received an apparently valid ticket, an itinerary of flights and a deceptive

cover story to be maintained during the security interview, and a ticketing record was created on the airline electronic database. Cover stories were unique and age- and gender-appropriate. For example, one participant (a male police officer from Hull) was told:

You have worked as a telecoms engineer for the past three years, in Northampton where you grew up. You are travelling to Chicago to take part in an international amateur fencing competition.

Participants were encouraged to undertake preparatory research to make their cover story convincing, by conducting internet searches to gather information relevant to the cover story. On the second and fifth days of the week prior to participation, each participant took part in a telephone interview with a researcher, in which they were asked to give a free verbal account of their cover story and the information that they had accessed to enrich their relevant story knowledge. A threshold for successful telephone interview was set, requiring participants to maintain at least two minutes of uninterrupted verbal account of their cover story, and to supply at least four pieces of information that were not included in the original cover story. All participants met this threshold in both interviews. On the day they attended the airport, they were told to dress appropriately for their destination and travel story, and to bring appropriate luggage and a current passport.

Screening interviews took place prior to check-in for 88 long-haul flights. On arrival at the airport, each mock passenger met a researcher who used a written protocol to explain participation. No information was provided about security procedures. Mock passengers were directed individually (out of sight of security staff) to join genuine passengers queuing to pass through either the suspicious signs or CCE security process, they were then interviewed by a security agent. If an agent was not

satisfied with the outcome of the interview (e.g., suspicious signs were detected but not resolved; the veracity of an account was not established), the passenger was identified for further screening under a separate mandatory process. Throughout the trial, security agents were fitted with digital audio recorders. Signs were posted informing passengers that security interviews may be recorded.

Results

Detecting Deception

Our principal interest was to determine which method was most effective for detecting deception (i.e., discriminating mock from real passengers). CCE agents identified 24 times more mock passengers (66%; 60% at month 1 and 72% at month 6) than suspicious signs agents (3%; 6% at month 1 and 0% at month 6), χ^2 (1) = 13.091, $p < .001$. Discrimination accuracy was assessed using a d' measure for each flight calculated from true positives (mock passengers) and false alarms (genuine passengers) identified as being deceptive. Agents using CCE, M_{CCE} = 3.016, 95% CI [2.65; 3.38], showed higher discrimination accuracy than agents using suspicious signs, $M_{\text{suspicious signs}}$ = 0.042, 95% CI [-.34; + .43], F (1, 136) = 119.853, p = < .001, η_p^2 = .47. The time of testing (Month 1 vs. 6) did not affect discrimination accuracy, all Fs < .912, all ps > .404.

Treating the genuine passengers who were identified as being deceptive as false alarms is a simplification, since it cannot be established whether they were mistakenly identified as deceptive or were actually being deceptive. However, base rates of identification of travelers identified as being deceptive in the six months prior to the trial (1 in 1,247 passengers) did not differ during the trial with either suspicious signs (1 in 1219 passengers), or CCE (1 in 1295 passengers), ps > 0.20.

Differences in Verbal Content

We hypothesized that the using CCE method would make it easier to detect deception because it generates more passenger verbalization than the suspicious signs method. To test this hypothesis, we analyzed the interview audio-recordings for the number of words spoken by each passenger, the number of words spoken by the agent, the number of questions asked by the agent, the types of questions asked (open, closed), and the temporal reference of questions (i.e., whether the question referred to an event in distant past, recent past, present, near future or far future). In addition, we counted the number of unique information items (i.e., information unknown to the agent and unavailable through documentation) given in the verbal account of each passenger. An independent coder coded 20% of the 404 transcripts. Inter-coder reliability for number and types of questions, screener and passenger words, and information items was high: r (82) = .827, p < .001; r (82) = .901, p < .001; r (82) = .988, p < .001; r (82) = .899, p < .001; r (82) = .841, p < .001, respectively.

One possibility is that agents treated mock passengers differently from real passengers from the outset, irrespective of the interview method being used (i.e., agents recognised mock passengers as being different from real passengers, so changed their interview approach accordingly). As a manipulation check, between-subjects Anovas were conducted for each of our measures that included Passenger Type (mock versus real) as a factor. These analyses revealed no significant main effects or interactions involving Passenger type for numbers of agent questions or words, all Fs < .009 and all ps > .507, passenger words or information items, Fs < .179, all ps > .673, or open, closed question types and temporalities, Fs < .179, all ps > .308. It appears that, although agents using CCE were able to detect mock

passengers while agents using suspicious signs could not, the way in which agents applied each interview method did not differ between mock and real passengers.

Examination of the frequencies of verbal behaviors (shown in Figure 1), irrespective of passenger type, indicates differences between CCE and suspicious signs interviews. Passengers (both mock and real) screened using CCE uttered more words, CI 95% [254.54; 275.40], than those screened using suspicious signs, CI 95% [58.39; 79.25], F (1, 400) = 683.543, $p < .001$, $\eta^2 = .63$. Passengers screened using CCE also revealed more information items, $M_{\text{CCE items}} = 11.64$, CI 95% [11.23; 12.05], than passengers screened using suspicious signs, $M_{\text{Suspicious Signs items}} = 0.76$, CI 95% [0.36; 1.17], F (1, 400) = 1379.924, $p < .001$, $\eta^2 = .77$. CCE agents uttered fewer words, CI 95% [118.08; 131.51], than suspicious signs agents, CI 95% [316.20; 329.63], F (1, 400) = 1686.806, $p < .001$, $\eta^2 = .80$. Thus, the results confirm our hypothesis that CCE yields more verbal behaviors from passengers.

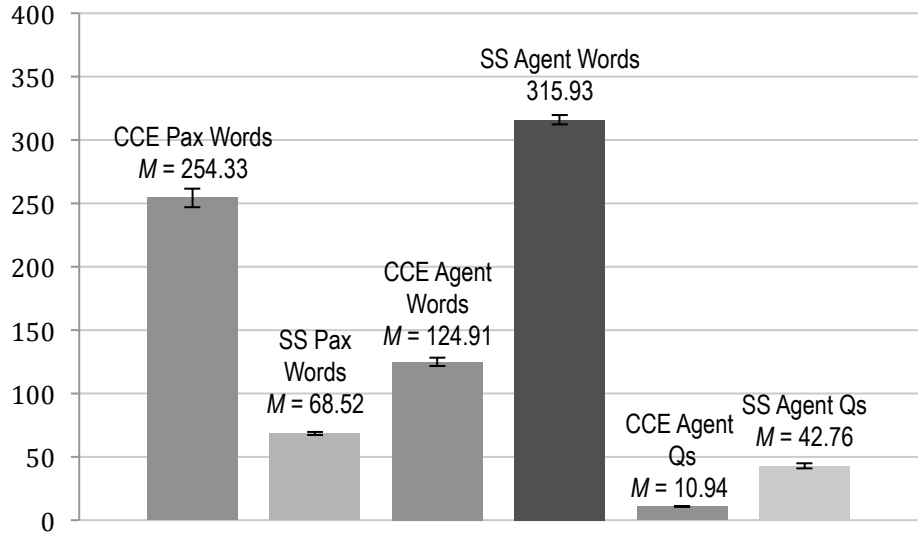

Fig. 1.

Mean number of words spoken by passengers (Pax) and agents (CCE and Suspicious Signs: SS) as a function of interview method (N =404).

Turning to the verbal behaviors of agents conducting the interviews, CCE agents asked fewer questions, CI 95% [10.35; 11.50], than suspicious signs agents, CI 95% [19.48; 20.64], F (1, 400) = 486.089, $p < .001$, $\eta^2 = .60$. CCE agents asked more open questions, CI 95% [3.06; 3.28], than suspicious signs agents, CI 95% [0.15; 0.37], F (1, 400) = 1376.880, $p < .001$, $\eta^2 = .76$, and their questions covered more temporal domains, CI 95% [2.10; 2.25], than those of suspicious signs questions, CI 95% [1.09; 1.24], F (1, 400) = 355.066, $p < .001$, $\eta^2 = .47$. However, CCE agents asked fewer closed questions, CI 95% [1.65; 2.04], F (1, 400) = 1645.126, $p < .001$, $\eta^2 = .80$, than suspicious signs agents, CI 95% [7.37; 7.76], (see Figure 2).

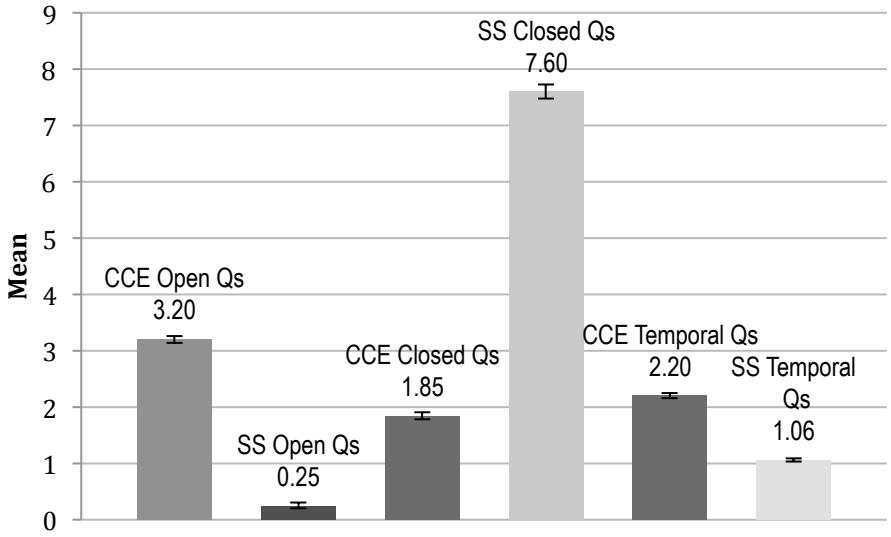

Fig. 2.

Mean question types asked by agents as a function of interview method (N =404).

An important practical consideration is how long it takes to administer each interview method. To examine this issue, we measured the duration of each interview (in seconds, from agent introduction until directing the passenger towards check-in). The difference in duration between CCE interviews ($M_{\text{CCE duration}}$ = 193.62, SD =

30.90) and suspicious signs interviews ($M_{\text{suspicious signs duration}}$ = 186.95, SD = 35.30), was not significant, F = 2.064, p =.152. As with other measures, no effects of Passenger Type (mock versus real) were found for interview duration, all Fs < 3.278, all ps > .071.

Change in Verbal Content

We also hypothesised that CCE is effective because it promotes tests of the veracity of passenger accounts, and these tests yield changes in the verbal behaviours of deceivers during an interview that can be used by security agents to discriminate them from truthful individuals. To test this hypothesis, we compared the number of words given by mock and genuine passengers in answer to the first and last open questions asked by a CCE security agent during an interview (i.e., comparing verbalisations before accounts have been challenged to verbalisations once deceptive individuals may have become aware that their accounts are being assessed). The number of words generated overall by mock passengers, $M_{\text{mock words}}$ = 24.80, CI 95% [23.43; 26.27], and genuine passengers, $M_{\text{genuine words}}$ = 26.54, CI 95% [25.1; 27.9] did not differ significantly, F (1, 199) = 2.96, p =0.087, η^2 = .015. The number of words uttered in answer to the first open question, $M_{\text{first words}}$ = 28.56, CI 95% [27.10; 29.95] was significantly larger than the number uttered in answer to the last open question, $M_{\text{last words}}$ = 22.80, CI 95% [21.85; 23.81], F (1, 199) = 52.90, p < .001, η^2 = .21. Importantly, the interaction between Passenger Type and Question was significant, F (1, 199) = 74.341, p < .001, η^2 = .27. Figure 3 illustrates the interaction, and shows that, while mock and genuine passengers uttered similar numbers of words in answers to the first open question of the interview, mock passengers uttered fewer words than genuine passengers in in answers to the last open question. Similarly, we predicted that mock passengers would include fewer information items than genuine passengers

in their answers to the focussed questions used by agents to test the veracity of accounts. The number of information items generated overall by mock passengers, $M_{\text{mock items}} = 3.43$, CI 95% [3.19; 3.67], was lower than by genuine passengers, $M_{\text{genuine items}} = 4.30$, CI 95% [4.05; 4.54], F (1, 199) = 24.7, $p < .001$, $\eta^2 = .11$.

Also, the number of information items yielded in the first test cycle, $M_{\text{first items}} = 4.49$, CI 95% [4.24; 4.74] was significantly larger than the number yielded in the last test cycle, $M_{\text{last items}} = 3.24$, CI 95% [3.07; 3.41], F (1, 199) = 87.056, $p < .001$, $\eta^2 = .30$. Again, the interaction between Passenger Type and Question was significant, F (1, 199) = 64.925, $p < .001$, $\eta^2 = .25$. Figure 3 illustrates this interaction, and shows that, while mock and genuine passengers included similar numbers of information items in their answers to the first test cycle of the interview, mock passengers gave fewer items than genuine passengers in their answers to the last test cycle.

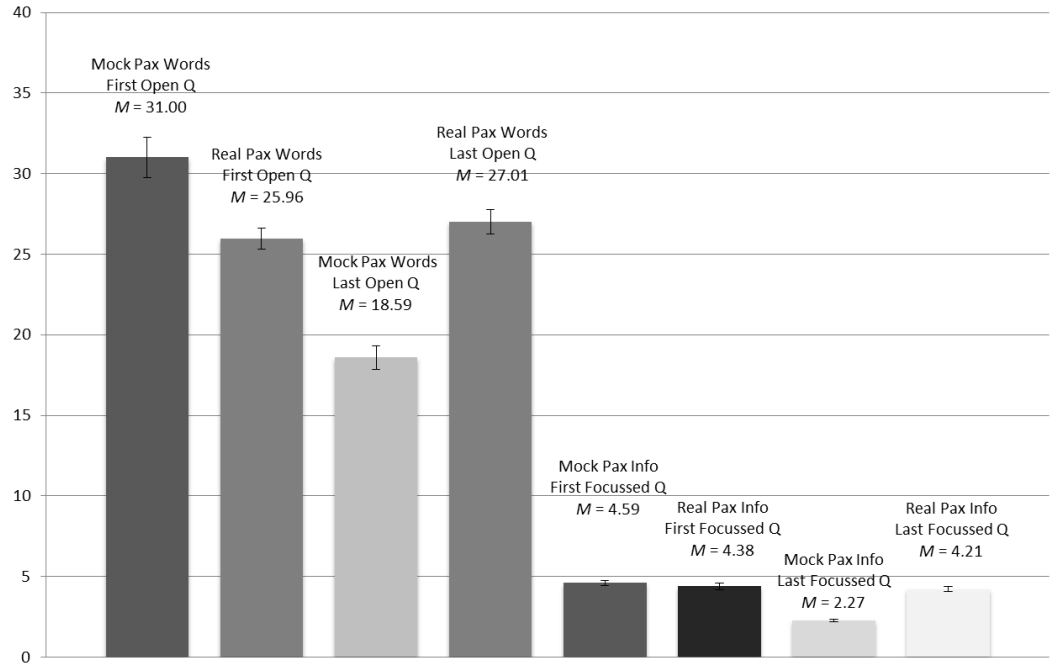

Fig. 3.

Mean number of words and information items as a function of passenger type (mock; real) and question position (first; last: N =404).

Discussion

The results of the field trial show a clear advantage for veracity testing over suspicious signs as a method for aviation security screening. CCE detected more mock passengers without increasing false alarm rates. With both mock and genuine passengers, CCE interviews yielded more passenger talk and information. At the same time, security agents produced less talk and asked fewer but more effective questions. The use of CCE changed the verbal behaviour of deceptive passengers, whose answers became shorter and had less information content by the end of the interview, while the answers of genuine passengers did not change. By using an information-gathering approach, first asking open questions about unpredictable topics that vary in their temporal reference, followed by test questions that seek information an individual should possess if they are being truthful, it is likely that CCE minimised cognitive demand for legitimate passengers but increased it for deceivers (Beckman, 2010).

The failure of a suspicious signs approach to detect mock passengers is consistent with the poor performance of behavioural indicators found in laboratory studies of deception (see de Paulo, 3003; Bond & de Paulo, 2006), and extends this finding to a composite method (where more than one indicator is being sought) that is tested in a field setting. The difference between suspicious signs and veracity testing approaches parallels a distinction in the decision-making literature between System 1 and System 2 modes of thinking (e.g., Evans, 2008; Kahneman, 2011). System 1 thinking uses cues in a task environment to trigger decision heuristics (e.g., Klein, 2004). System 2 thinking invokes more deliberative analytic decision-making and searches for counter-examples to initial inferences (e.g., Johnson-Laird, 2006). Security screening using the suspicious signs approach is analogous to System 1

thinking in using behavioural indicators to guide decision-making. Recognition of environmental cues and consequent retrieval of appropriate action sequences is a hallmark of expert decision-making, particularly in dynamic and time-critical domains (Klein, 2004; Schraagen, Militello, Ormerod, & Lipshitz, 2008). However, this kind of System 1 expertise develops from repeated exposure to cues in a relevant task environment. Security agents are rarely exposed to known incidents of deception, and cannot develop the kinds of automated expertise in cue recognition seen in other expertise domains. As a consequence, behavioural indicator methods for security screening necessarily comprise a rigid procedure in which the kinds of indicators to look for are prescribed and trained. The scripted nature of a suspicious signs interview makes it difficult to employ psychologically validated techniques such as tactical use of evidence (Dando et al., 2012), tests of expected knowledge (Blair et al., 2010), and unexpected questions (Vrij et al, 2009). Intuitive processing in deception detection of the kind promoted by a suspicious signs approach exacerbates truth and lie biases in deception judgements (Meissner & Kassin, 2002). Although the suspicious signs method is only one way of implementing behaviour detection, we argue that any method that relies on behavioural indicators will yield disappointing results, because individual behavioural cues are weakly correlated with deception.

Veracity testing is more effective than suspicious signs screening because it encourages a System 2 mode of thinking, in which the consistency of an individual's account is assessed analytically. Veracity testing directly addresses the problem identified in Porter, Gustaw and ten Brinke's (2010) Dangerous Decisions theory. CCE leads security agents to assess the content of individual's account, rather than relying on intuitive judgements about passenger appearance and disposition. Every

passenger is different, and CCE enables agents to adapt their interviewing to reflect these differences.

There are additional benefits of a veracity testing approach. Behavioural indicators associated with previous terrorist events may not predict future events; CCE identifies deceit in real time, allowing discovery of new kinds of threat. The unpredictability of CCE questioning breaks any lie script a deceiver might prepare (Von Hippel & Trivers, 2011), which reduces opportunities for reverse engineering of a security screening method and subsequent evasion by perpetrators (Chakrabarti & Strauss, 2002). In contrast, the suspicious signs method comprises a fixed sequence of closed questions to which responses can be rehearsed (e.g., during 'dry runs'). Finally, passive observation of passenger behaviours carries a risk of selective profiling that may disadvantage some ethnic, gender and age groups. CCE is applied equally to all passengers, avoiding inappropriate profile-based biases.

In closing, we note that the sensitivity of CCE for detecting deception is unique. Most studies employ base rates of around 50:50 deceivers to truth-tellers. Here, high rates of deception detection were obtained with a base rate of less than 1:1000 mock to genuine passengers. Thus, a more positive picture emerges of the contribution that can be made by psychological research to the protection of public safety than previously thought.

References

Beckman, J. F. (2010). Taming a beast of burden – On some issues with the conceptualization and operationalization of cognitive load. *Learning and Instruction,* **20***,* 250-264.

Blair, J. P., Levine, T. R., & Shaw, A.S. (2010). Content in context improves deception detection accuracy. *Human Communication Research,* **36,** 423-442.

Bond, C. F. & de Paulo, B. M. (2006). Accuracy of decision judgments. *Personality and Social Psychology Review,* **10,** 214-234.

British Security Industry Association (2008). Detecting behaviour to prevent aviation attacks. http://www.bsia.co.uk/aboutbsia/news/newsarticle/N4BCQB63655). [The easiest access to this source is by Internet].

Chakrabarti, S., & Strauss, A. (2002). Carnival booth: An algorithm for defeating the computer-based passenger screening system. *First Monday*, **10**, doi: http://dx.doi.org/10.5210%2Ffm.v7i10.992.

Dando. C. J., & Bull, R. (2011). Maximising opportunities to detect verbal deception: Training police officers to interview tactically. *Journal of Investigative Psychology and Offender Profiling, 8,* 189-202.

Dando, C. J., Bull, R., Ormerod, T. C., & Sandham, A. (2013). Helping to sort the liars from the truth-tellers: The gradual revelation of information during investigative interviews. *Legal and Criminological Psychology.* Article first published online: 20 APR 2013 DOI: 10.1111/lcrp.12016

de Paulo, B. M., Lindsay, J. J., Malone, B. E., Muhlenbruck, L., Charlton, K., & Cooper, H. (2003). Cues to deception. *Psychological Bulletin,* 129, 74-118.

Ekman, P. (2009). Lie-catching and micro-expressions. In C. Martin (Ed.). *The philosophy of deception*, Oxford: Oxford University Press.

Evans, J.S.B.T. (2008). Dual-processes accounts of reasoning. *Annual Review of Psychology*, 59, 255–278.

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

Johnson-Laird, P.N. (2006) *How We Reason*. Oxford: Oxford University Press.

Kahneman, D. (2011). *Thinking, Fast and Slow*. Penguin, ISBN: 9780141918921.

Klein, G. (2004). *The Power of Intuition: How to Use Your Gut Feelings to Make Better Decisions at Work.* Currency, ISBN 0-385-50289-3.

Levine, T. R. (2010). A few transparent liars: Explaining 54% accuracy in deception detection experiments. *Communication Yearbook, 34 (pp 40-61). Sage*.

Levine, T. R., Kim, R. K., & Blair, J. P. (2010). (In)accuracy at detecting true and false confessions and denials: An initial test of a projected motive model of veracity judgments. *Human Communication Research*, 36, 81-101.

Levine, T. R., Shaw, A. & Shulman, H. C. (2010). Increasing Deception Detection Accuracy with Strategic questioning. *Human Communication Research,* **36,** 216-231.

Martonosi, S.E. & Barnett, A. (2006). How Effective Is Security Screening of Airline Passengers? *Interfaces*, 36, 545–555. doi 10.1287/inte.1060.0231

Meissner, C. A., & Kassin, S. (2002). He's guilty! Investigator bias in judgments of truth and deception. *Law and Human Behavior,* **26,** 469-480.

Milne, R. & Bull, R. (1999). *Investigative interviewing: Psychology and practice.* Wiley, West Sussex.

Oxburgh, G., & Dando, C. J. (2011). Interviewing Witnesses and suspects: Where now in our search for the Truth? *British Journal of Forensic Practice. 13,* 135-147.

Oxburgh, G. E., Myklebust, T., Grant, T. (2010). The question of question types in police interviews: A review of the literature from a psychological and linguistic perspective. *International Journal of Speech, Language & the Law* 17, 45-66.

Porter, S., Gustaw, C., & ten Brinke, L. (2010). Dangerous decisions: The impact of first impressions of trustworthiness on assimilation of legal evidence and decisions of guilt. *Psychology, Crime and Law, 16,* 477-491. doi:10.1348/135532508X281520

Reddick, S. R. (2004). Point: The case for profiling. *International Social Science Review*, 79, 154-156.

Ross, L., & Nisbett, R. E. (1991). *The person and the situation: Perspectives of social psychology.* New York: McGraw-Hill.

Schraagen, J. M. C., Militello, L., Ormerod, T. C., & Lipshitz, R. (Eds) (2008). *Macrocognition and Naturalistic Decision Making.* Aldershot, UK: Ashgate Publishing Limited.

Shepherd, E. (2007). *Investigative Interviewing: The conversation management approach*. Oxford: Oxford University Press.

Taylor, P. J., Dando C. J., Ormerod, T. C., Ball, L. J., Jenkins, M. C., Sandham, A, Menacere, T. (2013) Detecting insider threats through language change. *Law Hum Behav*, United States: 37 (4), pp. 267-275. doi: 10.1037/lhb0000032

United States Government Accountability Office (US GAO) (2011). *Aviation Security: TSA is taking steps to validate the science underlying its passenger behavior detection program, but efforts may not be comprehensive.* GAO-11-

146T.

Vrij A., Leal S., Granhag P. A., Mann S., Fisher R. P., Hillman J., et al. (2009). Outsmarting the liars: the benefits of asking unanticipated questions. *Law & Human Behavior*, 33, 159–166. Doi: 10.1007/s10979-008-9143-y

Walczyk, J.J., Igou, F.P., Dixon, A.P., & Tcholakian, T. (2013). Advancing Lie Detection by Inducing Cognitive Load on Liars: A Review of Relevant Theories and Techniques Guided by Lessons from Polygraph-Based Approaches. *Frontiers of Psychology*, 4, Published online Feb 1, 2013. doi: 10.3389/fpsyg.2013.00014.

Weinberger, S. (2010). Intent to deceive? *Nature,* 465. 412-415.

Acknowledgments: The data reported in this paper are available from the authors. The research was funded by Her Majesty's Government Communications Centre, UK