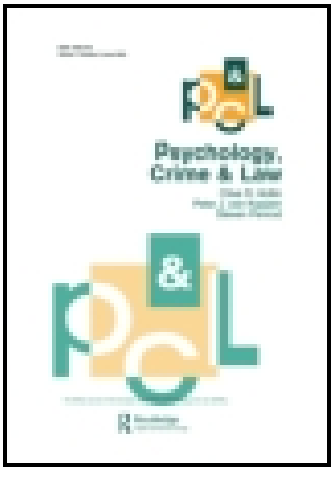

Psychology, Crime & Law
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/gpcl20

Detecting truth in suspect interviews: the effect of use of evidence (early and gradual) and time delay on Criteria-Based Content Analysis, Reality Monitoring and inconsistency within suspect statements

Alice Jennifer McDougall[a] & Ray Bull[a]
[a] School of Psychology, University of Leicester, Leicester, UK
Accepted author version posted online: 11 Dec 2014.Published online: 03 Jan 2015.

Psychology, Crime & Law, 2014
http://dx.doi.org/10.1080/1068316X.2014.994631

Routledge
Taylor & Francis Group

Detecting truth in suspect interviews: the effect of use of evidence (early and gradual) and time delay on Criteria-Based Content Analysis, Reality Monitoring and inconsistency within suspect statements

Alice Jennifer McDougall* and Ray Bull

School of Psychology, University of Leicester, Leicester, UK

(*Received 3 February 2014; accepted 21 November 2014*)

The strategic use of evidence in interviews with suspects has been shown to increase the ability of interviewers to accurately and consistently distinguish truthful from deceptive accounts. The present study considers the effect of early and gradual revelation of evidence by the interviewer, and the effect of shorter and longer delay on the verbal quality of truth-teller and liar statements within a mock crime paradigm. It was hypothesised that gradual disclosure of evidence (1) in terms of inconsistencies (a) within statements and (b) between statements and such evidence and (2) of the criteria of Criteria-Based Content Analysis (CBCA) and of Reality Monitoring (RM) would emphasise differences in the verbal quality of truth-teller and liar statements. Forty-two high school students took part in the study. The use of statement-evidence and within-statement inconsistency appears to be a robust cue to deception across interview style and delay. This indicates that gradual disclosure in interviews may increase interviewer accuracy in veracity decisions by eliciting statement inconsistencies. However, gradual revelation and delay affected the ability of CBCA and RM criteria to distinguish the veracity of suspect statements.

Keywords: interviewing to detect deception; verbal cues; within-statement consistency; statement-evidence inconsistency; interviewing strategies

Introduction

The ability to accurately and consistently distinguish truthful from deceptive accounts is of critical importance in many forensic settings. Although experience in interviewing has been shown to enhance police officers' accuracy in high-stakes situations (Mann, Vrij, & Bull, 2004), criminal justice investigators and lay people have often been found to be little better than chance at detecting deception (Akehurst, Köhnken, Vrij, & Bull, 1996; Vrij, 2008). This limited ability to accurately distinguish between truth and lies fits with a large body of research which indicates that cues to deception are typically found to be faint and unreliable (DePaulo et al., 2003; Sporer & Schwandt, 2006).

The development of the PEACE interviewing model (Milne & Bull, 1999) provided a structure for investigative interviewers to carry out effective and ethical interviews. This model also shifted the goal of interviewing, from obtaining a confession or disclosure to

*Corresponding author. Email: alicejmcd@gmail.com
This research was conducted at the University of Leicester.

eliciting as much accurate information as possible, so as to establish whether the statement is a truthful reflection of events. Nowadays, the majority of investigators in England who have been trained in the updated aspects of PEACE model use questioning to gradually disclose to suspects the relevant evidence/information rather than doing so at the beginning of the interviews, as was the case in the past (Bull, 2014; Moston, Stephenson, & Williamson, 1992; Walsh, Milne, & Bull, in submission). Recent research aims to find ways to elicit and amplify the cues available for detecting deception within these detailed accounts, so as to place detectors of deception in a more advantageous position (Vrij, Granhag, Mann, & Leal, 2011). The process of investigative interviewing and how evidence/information is revealed during interviews with suspects has latterly become a focus of research.

For suspects, research indicates that truth-tellers and liars enter interviews with different strategies and beliefs about how to convince the interviewer of their version of events (Hartwig, Granhag, Strömwall, & Doering, 2010; Hartwig et al., 2011; Strömwall & Willen, 2011). Typically liars formulate a lie script, planning their statement beforehand (Chan & Bull, 2014). Their use of avoidant strategies allows the deceiver to conserve information and keep their story simple, so as to remain flexible and avoid contradicting themselves (Hartwig et al., 2010). In contrast, innocent suspects have been found less likely to plan, aiming to appear spontaneous and unrehearsed (Hartwig, Granhag, & Strömwall, 2007). Instead they use forthcoming and cooperative verbal strategies, and try to actively address critical information in their statements (Hartwig et al., 2010).

It has been suggested that truth-tellers' approach comes from a subscription to 'truth beliefs' (Granhag & Hartwig, 2008; Hartwig, Granhag, Strömwall, & Kronkvist, 2006; Hartwig et al., 2007, 2010) and the 'illusion of transparency' where individuals feel that their subjective states are visible to others. Kassin (2005) emphasised that this illusion can put innocent suspects in a vulnerable position in interview settings, as claims of innocence can encourage confrontational interview styles.

During interviews with suspects, the interviewer needs to maximise the reliability of the information gained and manage this information in light of what evidence/information is already known (Gudjonsson, 2010; Milne & Bull, 1999). Evolving research indicates that certain styles of evidence/information disclosure may elicit cues to deceit and improve deception detection (Hartwig & Bond, 2011). Early disclosure (ED) of evidence/information is when the known relevant information is all revealed by the interviewer at the beginning of the interview. This means that the liar is aware of what information the interviewer knows, and can therefore try to provide an account that accommodates this. This can result in both liar and truth-teller statements that are consistent with the evidence, making veracity decisions difficult (Hartwig et al., 2006). In contrast, strategic disclosure of evidence can reveal inconsistencies between what a suspect has said and information that the interviewer knows but has not yet revealed.

A gathering body of research indicates that inconsistency discriminates truthful from deceptive statements (Clemens, Granhag, & Strömwall, 2011; Granhag, Strömwall, Willén, & Hartwig, 2013; Jordan, Hartwig, Wallace, Dawson, & Xhihani, 2012). Statement inconsistency can be divided into two types: statement-evidence inconsistency is the extent to which a suspect's statement is not consistent with the information known by the interviewer; whereas within-statement inconsistency refers to the extent to which a statement is not consistent within itself during an interview.

Hartwig and colleagues (2006) found that newly recruited trainee police officers trained in the strategic use of evidence (SUE) achieved far higher accuracy than those who interviewed suspects without any training. In that study SUE involved the interviewer revealing the evidence towards the end of the interview (late disclosure). In late disclosure interviews, truth-tellers' forthcoming strategies enable them largely to maintain consistency with the evidence, whereas liars' avoidant strategies require cognitive effort to conserve information whilst remaining credible and consistent. SUE exploits this difference in suspect strategies, facilitating statement-evidence inconsistencies that aid accurate veracity decisions (Hartwig et al., 2006, 2007; Jordan et al., 2012). Recently, however, it has been suggested that late disclosure may mean that more subtle within-statement inconsistencies are missed (Hartwig et al. 2011). This is because the interview style does not emphasise subtle differences in suspects' statements throughout the interview, for example through encouraging multiple accounts of the suspects' actions.

Research is now focusing on revealing the evidence gradually throughout the interview (Dando & Bull, 2011; Dando, Bull, Ormerod, & Sandham, 2014; Granhag et al., 2013). Gradual disclosure of evidence (GUE) differs from late disclosure, as the suspect is asked to describe the event and is then confronted with the evidence, a piece at a time, and asked to account for any discrepancies in their story. This approach highlights both statement-evidence and within-statement inconsistencies and therefore may provide more cues for detecting deception.

Dando and Bull (2011) found that a gradual disclosure technique resulted in more accurate veracity assessments than late and early disclosure. In this study the technique led to police interviewers citing the use of verbal cues for their decision-making, which other research has found to be a better guide to deception than non-verbal cues (DePaulo et al., 2003; Vrij, 2008). The type of verbal cues used by interviewers for veracity decisions were not explored by Dando and Bull, limiting our understanding of *how* gradual disclosure aids interviewer veracity decisions. In a further development of the SUE, the SUE incremental (SUE-I) also incorporates GUE (Granhag et al., 2013). SUE-I has been found to lead to lying suspects showing both more inconsistency within statements and between statement and evidence (Granhag et al., 2013). However, Sorochinski et al. (2014) more recently found late disclosure to provide more clues to deception.

Although such research has considered how SUE can affect the verbal cues of consistency, research has not yet considered how this may affect the diagnostic value of Criteria-Based Content Analysis (CBCA) and Reality Monitoring (RM) scores. CBCA and RM are two widely used techniques that examine verbal cues present in transcripts about real-life events to establish veracity. CBCA is based on the Undeutsch Hypothesis (Undeutsch, 1982), which states that a statement from memory differs in quality and content from a statement based on fabrication. For example, truthful accounts contain more sensory details and follow a more spontaneous structure than lies (Vrij & Mann, 2006). RM assumes that experienced events are processed and stored differently to events which are imagined (Johnson & Raye, 1981) so genuine memories are believed to contain more details, whilst imagined events are based on cognitive operations (inferences and suppositions). Although these techniques are intended for analysing real-life events, they have been applied to experimental settings to assess their usability (Vrij, 2008). Both perform well in laboratory settings, indicating that they can be considered useful as outcome measurements; however, it is important to consider that the consistency and

length of laboratory events flatter the scoring systems as they provide more criteria, in comparison to real-life events that may be shorter.

A further consideration is the effect of delay. Research has suggested that the process of providing credible truthful statements may become more difficult after a delay (Colwell, Hiscock-Anisman, Memon, Taylor, & Prewett, 2007; Hartwig et al., 2010). Memories for events somewhat distant in time would be expected to include fewer details than memories for very recent events, resulting in RM and CBCA criteria becoming less diagnostic of truth and lies (Blandón-Gitlin, Pezdek, Lindsay, & Hagen, 2009). Furthermore, normal forgetting may reduce the differences in storage and retrieval pathways that RM distinguishes between. For example, after a delay, rehearsal may lead 'external' memories to become more 'internal' and imagined events to become more 'external' (Vrij, 2008). Hartwig and colleagues (2006) hypothesised, but did not investigate, the possible effect of delay on statement-evidence inconsistency, suggesting that truth-tellers may become less consistent over time and therefore more similar to liars.

The present study aims to advance research in this field by looking at how the verbal content of truth-teller and liar statements may be affected by the timing of evidence disclosure within interviews (ED vs. GUE) and by delay in interviewing (shorter vs. longer):

It is hypothesised that:

(H_i) liars' statements will show significantly more statement-evidence inconsistency than will truth-tellers' statements,

(H_{ii}) GUE will be associated with more statement-evidence inconsistency in liars' statements than will the ED,

(H_{iii}) liars' statements will show significantly more within-statement inconsistency than will truth-tellers' statements,

(H_{iv}) GUE will encourage more within-statement inconsistency in liars' statements than will the ED,

(H_v) there will be an effect of delay on within-statement and statement-evidence inconsistency,

(H_{vi}) truth-tellers' statements will achieve significantly higher RM and CBCA total scores than will those of liars',

(H_{vii}) truth-tellers' statements will achieve significantly higher RM and CBCA total scores in the GUE interview than the ED,

(H_{viii}) truth-tellers' and liars' statements will achieve significantly lower RM and CBCA total scores after a longer delay than after a short delay,

(H_{ix}) liars will report using avoidant strategies whereas truth-tellers will report forthcoming strategies.

Method

The study consisted of three phases. In the first, participants either took part in mock-theft of exam answers from a teacher's office, or an innocent activity of retrieving an email address from the same office. In the second phase, participants were interviewed according to one of four interview conditions (ED or GUE with longer or shorter delay).

All participants were instructed to present themselves as innocent during this phase. Following this a variety of information was collected via questionnaire in the third phase.

Phase 1

Participants

The participants who took part in the study were all students at Rugby High School. They were quasi-randomly allocated by their participant number, using counterbalancing, to one of the conditions. Of the 42 participants, 70% were female and all were aged 16 or 17. 90% identified as white British, 2.5% identified as Asian, 7.5% identified as 'Other'.

A G-power calculation (Faul, Erdfelder, Lang, & Buchner, 2007) indicated the optimum number of participants as 48 with an effect size of .05; however, the number (and gender) of participants was constrained by limits imposed by the school.

Procedure

An event was developed, similar to the mock crime paradigm used by Hartwig and colleagues (2006, 2007), where participants actively take part in a role-play scenario. The rationale of this was to create an ecologically valid and memorable demonstration of deceptive/truthful behaviour. Two research assistants helped to administer parts of the study to ensure that the interviewer was blind to participant conditions (truth-teller or liar) and the participants did not become familiar with the interviewer.

The participants were randomly allocated to one of the two veracity conditions: truth-teller or liar. Participants in the truthful condition were instructed that they had permission to go into the teacher's office, find an email address in the address book, copy it down and then distribute it to the class by posting the note in a class post-box. Participants in the liar condition were instructed to go into a teacher's office, read the exam answers, copy them down to create a 'cheat sheet' and then distribute it to the class by posting the note in a class post-box.

These activities were designed to contain similar sub-events so that several pieces of evidence could be used in the interview allowing full operationalisation of GUE. This was to address the limitation of studies that only confront participants with a few pieces of evidence (Granhag et al., 2013; Hartwig et al., 2006, 2007; Jordan et al., 2012).

The evidence used was that: (1) the suspect did exceptionally well on the (fictitious) exam; (2) they were seen by a member of staff going into the office; (3) the notepad in the office had been used; (4) they were seen by another student rushing; (5) they were seen leaving the office with a piece of paper; (6) they were seen by a student near the class post-box; (7) a 'cheat sheet' on the office notepaper was found in the class post-box.

Phase 2

Pre interview instruction

After the event, participants in the longer delay condition booked an appointment to return to the school's Psychology Department to be interviewed in 7–10 days' time. Participants allocated to the shorter delay condition and the returning delay participants were given an interview preparation sheet. This provided instructions for the interview and offered participants 10 minutes within which they could plan before the interview.

Before being interviewed, all participants were instructed that there was probably some information suggesting that they could be guilty of stealing the exam paper answers, and this was why they were being interviewed. They were instructed that they needed to convince the interviewer that they were telling the truth about their actions. Thus, in the truth-teller condition participants were instructed to tell the truth, whereas in the liar condition participants were instructed to lie about their incriminating actions.

Interviews

The structure of the scripted interview schedules was based on the guidelines of best practice outlined in the PEACE Model (Milne & Bull, 1999). Free recall encouraged full and detailed accounts, specific questions incorporated the available evidence to examine statement-evidence inconsistency, and a second free recall phase intended to highlight within-statement inconsistency. Open ended questions and free recall were used as research indicates that this style of questioning is the best way to obtain detailed accounts (Oxburgh, Myklebust, & Grant, 2010; Vrij, Mann, Kristen, & Fisher, 2007). The techniques of evidence disclosure and directive questioning were designed to replicate the interviews used in Dando and Bull (2011).

Participants were interviewed by the first author and these interviews were audio recorded. Participants randomly allocated to be interviewed with the ED or GUE schedule. To ensure an objective approach by the interviewer, effort was taken to ensure the schedules were highly structured. This was achieved with questions that mirrored the purpose, or were the same, across the GUE and ED schedules. All interviews began with an introduction and short explanation of the purpose of the interview.

Early evidence disclosure (ED)

The ED interview began with an 'Evidence Disclosure Phase' in which the interviewer listed all of the seven items of evidence against the suspect. The suspect was instructed not to respond during this stage. A 'Free Recall Phase' followed, in which the participant/suspect was asked to describe their actions in as much detail as possible. Participants were then asked to account for each piece of evidence in turn during the 'Questioning Phase' and were challenged to clarify any contradictions.

Gradual use of evidence (GUE)

The GUE interview began with a 'Free Recall Phase' where the suspect was asked to describe their actions on the day (relating to the event) in as much detail as possible. This was followed by a 'Questioning Phase' where the interviewer asked the suspect questions related to each of the seven pieces of evidence, for example, in relation to the evidence that the suspect had been seen going into the office before the exam, the question probe 'What did you do before the exam?' was used. The suspects answers were fully explored with further probes and any contradictions with the evidence were challenged. The interviewer did not fully disclose the nature of the evidence until the suspect had provided a full answer.

The 'Questioning Phase' in both interviews was followed by a second 'Free Recall Phase' and the interviews ended with a closure section. In both interviews, standardised prompts were used to encourage detailed accounts, such as 'Can you tell me more about that?' 'Are you sure you have mentioned everything?'

Post-interview questionnaire

After the interview was completed participants were asked to return to the research assistant and fill in a short questionnaire. Participants completed a questionnaire asking to what extent they told the truth or lied within the interview and how motivated they were during the interview. All were rated on a five-point Likert scale (1: low, 5: high). Participants were also asked if they had a strategy during the interview. If they answered yes they were asked to describe their strategy, if they answered no, they were asked why they chose to have no strategy. Similar questions regarding interview experience have been used in several prior studies (such as Hartwig et al., 2007; Strömwall, Hartwig, & Granhag, 2006).

Phase 3

Each interview was transcribed verbatim and then rated using the coding sheets by the first author, to obtain CBCA and RM score and frequency totals and within-statement and statement-evidence inconsistency scores.

Coding of transcripts

The operationalisations of CBCA criteria and RM criteria outlined in Vrij (2008) were used (two CBCA criteria were omitted as they are specific to analysing children's allegations of sexual abuse).

The following scores and frequencies were used to record the number of times each of the criteria occurred in the transcripts:

(a) scores were recorded on a three-point Likert scale: criterion absent (1); criterion present (2); criterion strongly present (3), this is the method traditionally used in coding CBCA and RM criteria,
(b) frequencies, recording every instance of present criteria, were intended to be more sensitive to differences, and have been used in previous research instead of scores for this reason (Vrij et al., 2007).

The scores and frequencies were then summed to create score totals and frequency totals. For RM totals, the criterion of cognitive operations is negatively scored as it indicates deception and so this item was subtracted from the total of the rest of the RM criteria.

The two types of statement inconsistency were scored on three-point Likert scales. Within-statement inconsistency was scored as: perfectly consistent (1); possibly consistent (2); inconsistent (3). Statement-evidence inconsistency was scored as: completely consistent with evidence (1); evidence not mentioned or possibly consistent (2); evidence contradicted by statement (3). Use of these measures appears to be gaining popularity, with similar scales used in recent research (Granhag et al., 2013; Hartwig et al., 2010; Jordan et al., 2012).

Limitations of time and resources prevented the use of interrater reliability measures. Although this is obviously a limitation of the study, it is reflective of in vitro use of the measures (Vrij, 2005) and care was taken to review and apply details from other studies management of the scales and scoring (Granhag, Strömwall, & Landström, 2006; Vrij & Mann, 2006).

Results

An exploration of the data indicated that, because of deviations from normality, the data did not meet assumptions of parametric statistical testing and therefore non parametric tests were used.

Manipulation check

Participants in the truth-teller condition reported telling the truth to a significantly higher extent (Mdn = 1.00) than did lying participants (Mdn = 5.00), U = 4.00, $p < .001$, $r = –.88$.

Participants were highly motivated during the task ($M = 2.03$, SD = .80) and there was no significant difference between liars and truth-tellers, (Mdn = 2.00, 2.00), U = 145.00, $p = .10$, $r = –.26$.

A potential confound for this study is that length of account has been suggested to affect CBCA and RM analyses, with longer accounts more likely to receive higher scores than shorter narratives (Blandón-Gitlin et al., 2009). Interviews lasted from 3.47 to 9.44 minutes. A Kruskal–Wallis test found no significant effect on statement length of truth-teller, liar, delay or interview type, H(7) = 8.99, $p = .25$.

Statement-evidence inconsistency

As mentioned above, suspects were confronted with seven items of evidence during the interviews. The numerical references to these pieces of evidence were outlined in the Method section.

Kruskal–Wallis tests across four conditions for the effect of interview style (ED vs. GUE by Liar vs. Truth-teller) revealed a significant effect for statement-evidence consistency regarding evidence item 3, H(3) = 13.21, $p = .004$, evidence item 5, H(3) = 14.17, $p = .003$ and evidence item 6, H(3) = 15.66, $p = .001$.

Further Kruskal–Wallis tests across four conditions for the effect of delay (Longer vs. Shorter by Liar vs. Truth-teller) revealed a significant effect for statement-evidence inconsistency regarding evidence item 3, H(3) = 21.75, $p < .001$, evidence item 5, H(3) = 20.24, $p < .001$ and evidence item 6, H(3) = 23.67, $p < .001$.

Follow-up Mann–Whitney U tests were then conducted where appropriate. As shown in Table 1, in both the ED and GUE interviews liars showed higher statement-evidence inconsistency than truth-tellers, supporting H_i. However, neither liars nor truth-tellers showed differences across interview style, with the gradual interview style not encouraging greater statement-evidence inconsistency as predicted (H_{ii}). In the longer delay, liars showed higher statement-evidence inconsistency than truth-tellers but, in the shorter delay condition there was no significant difference. Truth-tellers showed no difference in statement-evidence inconsistency across delay condition, but liars demonstrated higher statement-evidence inconsistency after a longer delay than after a shorter delay, supporting H_v. (Evidence item 5 was not significant due to the Bonferroni correction, setting a 0.125 level for significance.)

Within-statement inconsistency

A Kruskal–Wallis test across four conditions for the effect of interview style (ED vs. GUE by Liar vs. Truth-teller) revealed a significant effect for within-statement inconsistency, H(3) = 22.96, $p < .001$.

Table 1. The effect of interview style and delay on statement-evidence inconsistency.

	Truth-teller			Liar								
	Median			Median			Significance (p) and effect size (r)					
	Evidence 3	Evidence 5	Evidence 6	Evidence 3	Evidence 5	Evidence 6	Evidence 3		Evidence 5		Evidence 6	
Interview style							p	r	p	r	p	r
ED	1.00	1.00	1.00	3.00	3.00	2.00	.005*	–.58	.005*	–.58	.001*	–.69
GUE	1.00	1.00	1.00	2.00	3.00	2.00	.007*	–.55	<.004*	–.59	.0095*	–.52
Significance (p)	.29	.29	.14	.34	.43	.37						
Effect size (r)	–.12	–.12	–.24	–.092	–.38	–.075						
Delay												
Shorter	1.00	1.00	1.00	1.50	1.50	2.00	.25	–.15	.17	–.21	.19	–.19
Longer	1.00	1.00	1.00	3.00	3.00	2.50	<.001*	–.94	<.001*	–.94	<.001*	–.92
Significance (p)	.034	.034	.015	.002*	.016**	.0015*						
Effect size (r)	–.041	–.041	–.49	–.64	.48	–.66						

*Significant effect.
**Not significant due to Bonferroni correction.

Table 2. The effect of interview style and delay on within-statement inconsistency.

	Truth-teller	Liar	Significance and effect size	
	Median	Median	*p*	*r*
Interview style				
ED	1.00	1.50	.027	–.43
GUE	1.00	3.00	<.001*	–.85
Significance (*p*)	.16**	.007*		
Effect size (*r*)	–.22	–.55		
Delay				
Shorter	1.00	1.00	.015**	–.50
Longer	1.00	2.50	<.001*	–.86
Significance (*p*)	.16**	.008*		
Effect size (*r*)	–.22	–.54		

*Significant effect.
**Not significant due to Bonferroni correction.

A further Kruskal–Wallis test across four conditions for the effect of delay (Longer vs. Shorter by Liar vs. Truth-teller) revealed a significant effect for within-statement inconsistency, $H(3) = 24.96$, $p < .001$.

Follow-up Mann–Whitney U tests were then conducted where appropriate. As shown in Table 2, with GUE liars showed more within-statement inconsistency than truth-tellers, but with ED there was no difference. This is contrary to H_{iii} that predicted an overall difference for liars across interview type, this demonstrated the advantage that knowing the evidence can have when attempting to appear consistent. There was no difference for within-statement inconsistency of truth-tellers across interview style, but liars were more inconsistent in the GUE than the ED supporting H_{iv}. In the longer delay condition liars displayed more within-statement inconsistency than truth-tellers, but in the shorter delay condition there was no difference. Truth-tellers showed no difference across delay condition, but liars displayed more within-statement inconsistency in the longer delay than the shorter delay. These results provide further support for H_v.

CBCA and RM totals

Mann–Whitney U tests revealed that truth-teller statements were significantly higher than liars in CBCA frequency (Mdn = 31.00, 22.50), $U = 132.50$, $p < .05$, $r = –.29$ and CBCA scores (Mdn = 9.50, 7.00), $U = 134.00$, $p < .05$, $r = –.29$, RM frequency (Mdn = 42.50, 28.00), $U = 74.5$, $p < .001$, $r = –.54$ and RM scores (Mdn = 10.50, 8.00), $U = 97.5$, $p < .001$, $r = –.44$. This supports H_{vi} and is in line with the notions underlying CBCA (Undeutsch, 1982) and RM (Johnson & Raye, 1981).

Kruskal–Wallis tests across four conditions for the effect of interview style (ED vs. GUE by Liar vs. Truth-teller) revealed a significant effect for RM frequency, $H(3) = 13.87$, $p = .003$, and RM scores, $H(3) = 8.01$, $p = .056$, but no significant effect for CBCA frequency $H(3) = 3.77$, $p = .29$, or CBCA scores $H(3) = 3.49$, $p = .32$.

Further Kruskal–Wallis tests across four conditions for the effect of delay (Longer vs. Shorter by Liar vs. Truth-teller) revealed a significant effect for RM frequency,

Table 3. The effect of interview style and delay on RM score and frequency totals.

	Truth-teller		Liar					
	Median		Median		Significance (p) and Effect Size (r)			
					Score		Frequency	
Interview style	Score	Frequency	Score	Frequency	*p*	*r*	*p*	*r*
ED	10.50	42.50	8.00	25.50	.013*	–.50	.001*	–.72
GUE	10.50	41.50	8.50	32.00	.40	–.39	.061	–.35
Significance (*p*)	.41	.50	.35	.07				
Effect size (*r*)	–.05	–.03	–.09	–.33				
Delay								
Shorter	11.00	49.50	8.00	28.00	<.001*	–.75	<.001*	–.71
Longer	8.50	36.00	8.50	28.50	.29	–.12	.06	–.35
Significance	.021**	.012	.26	.47				
Effect size	–.45	–.51	–.15	–.02				

*Significant effect.
**Not significant due to Bonferroni correction.

H(3) = 14.89, p = .002, and RM scores, H(3) = 13.41, p = .004 but no significant effect for CBCA frequency, H(3) = 5.88, p = .12, or CBCA scores, H(3) = 5.97, p = .11.

Follow-up Mann–Whitney U tests were then conducted where appropriate. As shown in Table 3, in the ED truth-tellers' RM score and frequency totals were higher than liars', however in the GUE no differences were found, supporting H_{vi} for RM totals but not for CBCA totals. Neither truth-tellers or liars showed a difference across interview style, contrary to the prediction made in H_{vii}. In the shorter delay truth-tellers' RM score and their RM frequency totals were higher than liars, however in the longer delay no significant differences were found. In the longer delay condition truth-teller RM frequency totals were lower than in the short delay condition. There were no differences across delay conditions for liars' RM score or their RM frequency totals, providing support for H_{viii}, but only for RM totals. This reflects normal forgetting in truth-tellers resulting in statements made after a longer delay containing significantly lower RM totals than those made after a shorter delay (Blandón-Gitlin et al., 2009). Similarly to the findings of Vrij and colleagues (2009), liars' statements in the present study were similar across delay conditions.

Suspect strategy

Content analysis of the open-ended questions about strategy revealed several categories. Full presentation of this qualitative analysis of participants' strategies and beliefs is not possible given the constraints on the length of the current document, however the categories are summarised in Table 4, indicating support for H_{ix}.

Table 4 shows truth-tellers reported strategies that demonstrate use of approach strategies whilst liars reported use of avoidant strategies.

Table 5 shows that the motivations for having no strategy were similar for truth-tellers and liars except for truth-tellers citing 'innocence', reflecting endorsement of truth beliefs.

Table 4. A summary of categories of truth-teller and liar strategies used when the suspect chose to plan.

Truth-teller, plan	Liar, plan
Be calm	Avoid incriminating details
Be consistent	Be calm
Being an honest person	Be convincing
Firmly deny guilt	Be consistent
Tell the truth	Firmly deny guilt
	Implicate others
	Keep it simple
	Play the role of innocent

Discussion

RM and CBCA totals

As expected, frequency totals (for RM) produced more statistically significant differences than scores, due to the sensitivity of the former measure (Vrij et al., 2007). The calculated effect sizes also demonstrated that the significant effects were large (Cohen, 1992). This indicates that, frequency totals may have an advantage in uncovering significant effects, in comparison to the use of Likert scales that indicate the overall strength in presence of criteria. This process is however more time-consuming and sensitive, possibly limiting the wider application of RM in suspect interviews in field settings.

RM frequency totals were significantly higher for truth-tellers than liars, but only for ED. With ED, both truth-tellers and liars are told early on of the evidence against them. Because of this, truth-tellers may feel the need to provide more details in their statement as they are aware that merely accounting for the already revealed evidence is not enough to distinguish them from liars. With GUE, truth-tellers may offer direct, concise answers to evidence related questions, whilst liars on the other hand may give more details in their answers believing this will increase their credibility (Vrij, 2008), and provide the flexibility to incorporate any evidence when it is revealed.

Further to this, research has indicated that truth beliefs and certain lines of questioning can lead to short, abrupt answers in truth-tellers (Kassin, 2005). CBCA and RM analyses were developed for use with information gathering interviews and, although PEACE guidelines encourage this, it may be necessary to go further to obtain detailed recall. For example Colwell's ACID interview (2007), encourages truth-tellers to shine through with their flexible memory retrieval through the use of mnemonic techniques. Recent developments are also encouraging suspect interviewers to become more interactive

Table 5. A summary of categories of truth-teller and liar motivations for having no plan for the interview.

Truth-teller, no Plan	Liar, no Plan
Be spontaneous	Be spontaneous
Did not know what would happen	Did not know what would happen
Innocence	

and incorporate the purposeful use of evidence. This may disrupt the quality of truth-teller statements, as suspects may react differently to probes (questions around evidence) than to prompts (questions encouraging recall), possibly resulting in statements which score lower than expected in RM and CBCA totals. This study's finding is a strong indication that it is not the verbal cues assessed by RM that are relied upon for diagnostic decisions which underpin the improved accuracy of GUE (Dando & Bull, 2011; Dando et al., 2014).

Statement-evidence inconsistency

Across both interview styles, liars showed significantly more statement-evidence inconsistency than truth-tellers. These results are similar to those found by Granhag and colleagues (2013). The GUE did not promote greater statement-evidence inconsistency as predicted, possibly because suspects anticipated the evidence or because the evidence was not strong enough (Granhag et al., 2013). The results of the present study add to a gathering body of research that indicates statement-evidence inconsistency to be a robust cue to deception (Clemens et al., 2011; Hartwig et al., 2011; Jordan et al., 2012). The advantage is that the idea of consistency with evidence and the style of interviewing is likely to be quick to train. Evidently in the field, the types and strength of evidence available and the suspect's prior knowledge of this evidence, will greatly influence the operationalisation of GUE and usability of statement-evidence inconsistency as a diagnostic indicator. This may require the consideration of current evidence management and a proposed structure for *when* GUE should be used (Walsh et al., in submission).

Within-statement inconsistency

Within-statement consistency is less dependant on the specifics of the evidence presented. Content analysis found that both truth-tellers and liars planned to be consistent during their interviews. This is likely to be due to the belief that inconsistencies may lead their statement to be challenged and their credibility undermined (Clemens et al., 2011). For the GUE interview, liars exhibited significantly higher within-statement inconsistency than truth-tellers, similar to the findings of Granhag and colleagues (2013). GUE interviews encourage greater within-statement inconsistency in liar statements than ED, because the liars are asked to account for the evidence before it is revealed, resulting in changes within their statement as they try to accommodate and remain flexible during questioning. With ED there was no significant difference between liar and truth-teller within-statement inconsistency as the liars were able to incorporate the disclosed evidence into their statements and therefore provide consistent accounts. Truth-teller within-statement consistency remained high across interview style, probably because they are able to be more flexible in their memory retrieval (Colwell et al., 2007). So it would appear that within-statement consistency is the measure in which truth-tellers can best shine through. Through training with GUE interviewers could be taught how to elicit these cues that are straightforward to interpret.

Truth-teller strategies of directly accounting for evidence may result in high statement-evidence and within-statement consistency, but may also result in short direct accounts of actions which lack detail. This may partly be due to participants' endorsement of truth beliefs demonstrated by the questionnaire. It is possible that,

although GUE increases inconsistency in liar statements, it may cause a convergence of RM and CBCA scores in liar and truth-teller statements.

Delay

The present study found that truth-tellers did experience normal forgetting over the delay period that liars did not (as shown in RM totals), whilst they maintained high statement-evidence consistency and within-statement consistency.

The effect of delay on RM totals led to a convergence of the quality of liars' and truth-tellers' statements, therefore reducing the diagnostic value of RM. It is possible that, after a longer delay than used in the current study, truth-teller statements may begin to include the CBCA criteria of motivation-related contents, such as admitting lack of memory. These criteria are specific to truth-teller accounts (Vrij et al., 2004), but their presence is unlikely to outweigh the greater detail of deceptive accounts, reiterating previously voiced concerns regarding the weighting of CBCA criteria (Masip, Sporer, Garrido, & Herrero, 2005; Vrij, 2005), and again suggesting that it is not the quality of statements, measured by RM, that underlie the efficacy of GUE.

For liars, the longer delay exacerbated statement-evidence inconsistency and within-statement inconsistency enabling them to be distinguished from truth-tellers. This provides further support for the claim that statement-evidence and within-statement inconsistency are robust cues to deception (Granhag et al., 2013). The evidence indicates that normal forgetting is experienced by truth-tellers, leading to disruptions in statement quality, however the GUE allows truth-tellers to demonstrate their flexibility in retrieval, assessed by their consistency over time. This is an extremely positive indicator for the potential success for the use of GUE in the field.

Overall, these results demonstrate that interview structure, delay, and suspect strategies can have an effect on the quality of statements. Truth-teller statements may be more sensitive to these effects leading to lower-quality statements, inhibiting RM's ability to distinguish them from liar statements and in fact may be putting liars in an advantageous position. It is reasonable to assume that there would be even more of an impact of these (and other) factors in the field. The results do, however, indicate that GUE disrupts typical patterns of statement quality, encouraging liars to provide more details in their statements and therefore increasing the likelihood of inconsistencies being identified.

Limitations

Contrary to predictions, significant differences in CBCA total scores for truth-tellers and liars were not maintained when considering interview type and delay. This is unlikely to be due to small sample size or lack of statistical power as other significant effects have been uncovered in the analyses. It is proposed that limitations in methodology are more likely to be why no significant effect was found. Due to differences in operationalisation, there is no consensus about the training required to become competent at scoring statements for CBCA or RM criteria (Vrij, 2005). In the present study the coder was able to familiarise herself sufficiently with the criteria for RM, but the larger number of CBCA criteria and their more ambiguous interpretation may have had an impact on the data obtained. It is worth noting that to be practical for use in applied settings, a technique's scoring procedure needs to be accessible and quick to train.

A possible limitation of this study is that liars did not show significantly different statement-evidence inconsistency from truth-tellers for four of the seven evidence pieces. It could be argued that this may have limited the efficacy of the GUE interview as suspects were able to sufficiently account for half of the evidence. However the presence of evidence of varying strengths is more likely to be available in real-life interviews (Granhag & Hartwig, 2008). Furthermore, some prior research used just three evidence pieces (Hartwig et al., 2006, 2007; Jordan et al., 2012) and Granhag and colleagues (2013) only used one.

Using a mock crime paradigm is advantageous as participants engage with the active event, and this has been found to result in stronger support for RM (Masip et al., 2005) and CBCA (Vrij, 2008) criteria, however there are also limitations to this methodology (Gödert, Gamer, Rill, & Vossel, 2005). In the present study, the simplicity of the tasks may have limited the production of rich statements during interviews (Hartwig et al., 2011). Therefore the cognitive demand upon the suspects may not have been that high. The present study nevertheless attempted to increase complexity through including multiple sub-events. The interviews were also largely scripted which limited flexibility within interviews and did not mirror real life, but did make all interviews comparable.

It could be argued that the observed differences between delay conditions could be attributed to participants in the longer delay condition having more time to plan and prepare for their statement. To control for this, participants in both delay conditions were only given the interview preparation sheets 10 minutes prior to being interviewed.

The use of these students as participants could be a further limitation. All were attending a 'selective' grammar school. A more representative sample of participants would allow findings to be generalised and may include a larger percentage of people with average or below cognitive ability.

Conclusion

The results indicate that within-statement inconsistencies were significantly more present in liar statements when interviewed with GUE. It is encouraging that differences in truth-teller and liar statement-evidence and within-statement inconsistency remained after a longer delay, adding to research that indicates that these may be robust diagnostic cues. This study provides further support for GUE placing interviewers in an advantageous position in their interviews with suspects.

Acknowledgements

We would like to thank Rugby High school for kindly allowing us to conduct the research in their school, and for the enthusiasm of the students who were willing to give their time to participate.

Disclosure statement

The authors are aware of no conflicts of interest, financial or otherwise.

References

Akehurst, L., Köhnken, G., Vrij, A. & Bull, R. (1996). Lay persons' and police officers' beliefs regarding deceptive behaviour. *Applied Cognitive Psychology*, *10*, 461–471. doi:10.1002/(SICI)1099-0720(199612)10:6<461::AID-ACP413>3.0.CO;2-2

Blandón-Gitlin, I., Pezdek, K., Lindsay, D. S., & Hagen, L. (2009). Criteria-based content analysis of true and suggested accounts of events. *Applied Cognitive Psychology*, *23*, 901–917. doi:10.1002/acp.1504

Bull, R. (2014). When in interviews to disclose information to suspects and to challenge them? In R. Bull (Ed.), *Investigative interviewing*. New York, NY: Springer.

Chan, S., & Bull, R. (2014). The effect of co-offender planning on verbal deception. *Psychiatry, Psychology and Law*, *21*, 457–464. doi:10.1080/13218719.2013.835703

Clemens, F., Granhag, P. A., & Strömwall, L. A. (2011). Eliciting cues to false intent: A new application of strategic interviewing. *Law and Human Behavior*, *35*, 512–522. doi:10.1007/s10979-010-9258-9

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*, 155–159. doi:10.1037/0033-2909.112.1.155

Colwell, K.,Hiscock-Anisman, C. K.,Memon, A.,Taylor, L., & Prewett, J. (2007). Assessment Criteria Indicative of Deception (ACID): An integrated system of investigative interviewing and detecting deception. *Journal of Investigative Psychology and Offender Profiling*, *4*, 167–180. doi:10.1002/jip.73

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. doi:10.1002/jip.145

Dando, C. J., Bull, R., Ormerod, T. C., & Sandham, A. L. (2014). Helping to sort the liars from the truth-tellers: The gradual revelation of information during investigative interviews. *Legal and Criminological Psychology*. doi:10.1111/lcrp.12016

DePaulo, B. M., Lindsay, J. J., Malone, B. E., Muhlenbruck, L., Charlton, K., & Cooper, H. (2003). Cues to deception detection. *Psychological Bulletin*, *129*(1), 74–118. doi:10.1037/0033-2909.129.1.74

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*, 175–191. doi:10.3758/BF03193146

Gödert, H., Gamer, M., Rill, H.-G., & Vossel, G. (2005). Statement validity assessment: Inter-rater reliability of criteria-based content analysis in the mock-crime paradigm. *Legal and Criminological Psychology*, *10*, 225–245. doi:10.1348/135532505X52680&&&

Granhag, P. A., & Hartwig, M. (2008). A new theoretical perspective on deception detection: On the psychology of instrumental mind-reading. *Psychology, Crime & Law*, *14*, 189–200. doi:10.1080/10683160701645181

Granhag, P. A., Strömwall, L. A., & Landström, S. (2006). Children recalling an event repeatedly: Effects on RM and CBCA scores. *Legal and Criminological Psychology*, *11*(1), 81–98. doi:10.1348/135532505X49620

Granhag, P. A., Strömwall, L. A., Willén, R. M., & Hartwig, M. (2013). Eliciting cues to deception by tactical disclosure of evidence: The first test of the Evidence Framing Matrix. *Legal and Criminological Psychology*, *18*, 341–355. doi:10.1111/j.2044-8333.2012.02047.x

Gudjonsson, G. H. (2010). Psychological vulnerabilities during police interviews: Why are they important? *Legal and Criminological Psychology*, *15*, 161–175. doi:10.1348/135532510X500064

Hartwig, M., & Bond, C. F. (2011). Why do lie catchers fail? A lens model meta analysis of human lie judgements. *Psychological Bulletin*, *137*, 643–659. doi:10.1037/a0023589

Hartwig, M., Granhag, P. A., & Strömwall, L. A. (2007). Guilty and innocent suspects' strategies during police interrogations. *Psychology, Crime & Law*, *13*, 213–227. doi:10.1080/10683160600750264

Hartwig, M., Granhag, P., Strömwall, L., & Doering, N. (2010). Impression and information management: On the strategic self-regulation of innocent and guilty suspects. *The Open Criminology Journal*, *3*, 10–16.

Hartwig, M., Granhag, P. A., Strömwall, L. A. & Kronkvist, O. (2006). Strategic use of evidence during police interviews: When training to detect deception works. *Law and Human Behavior*, *30*, 603–619. doi:10.1007/s10979-006-9053-9

Hartwig, M., Granhag, P. A., Strömwall, L., Wolf, A. G., Vrij, A., & Roos af Hjelmsäter, E. (2011). Detecting deception in suspects: Verbal cues as a function of interview strategy. *Psychology, Crime & Law*, *17*, 643–656. doi:10.1080/10683160903446982

Johnson, M. K., & Raye, C. L. (1981). Reality monitoring. *Psychological Review*, *88*(1), 67–85. doi:10.1037/0033-295X.88.1.67

Jordan, S., Hartwig, M., Wallace, B., Dawson, E., & Xhihani, A. (2012). Early versus late disclosure of evidence: Effects on verbal cues to deception, confessions, and lie catchers' accuracy. *Journal of Investigative Psychology and Offender Profiling*, *9*(1), 1–12. doi:10.1002/jip.1350

Kassin, S. M. (2005). On the psychology of confessions: Does innocence put innocents at risk? *American Psychologist*, *60*, 215–228. doi:10.1037/0003-066X.60.3.215

Mann, S., Vrij, A., & Bull, R. (2004). Detecting true lies: Police officers' ability to detect suspects' lies. *Journal of Applied Psychology*, *89*(1), 137–149. doi:10.1037/0021-9010.89.1.137

Masip, J., Sporer, S. L., Garrido, E., & Herrero, C. (2005). The detection of deception with the reality monitoring approach: A review of the empirical evidence. *Psychology, Crime & Law*, *11*(1), 99–122. doi:10.1080/10683160410001726356

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

Moston, S., Stephenson, G., & Williamson, T. (1992). The effects of case characteristics on suspect behaviour during police questioning. *British Journal of Criminology*, *32*, 23–40.

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 and the Law*, *17*(1), 45–66. doi:10.1558/ijsll.v17i1.45

Sorochinski, M., Hartwig, M., Osborne, J., Wilkins, E., Marsh, J., Kazakov, D., & Granhag, P. (2014). Interviewing to detect deception: When to disclose the evidence? *Journal of Police and Criminal Psychology*, *29*(2), 87–94. doi:10.1007/s11896-013-9121-2

Sporer, S. L., & Schwandt, B. (2006). Paraverbal indicators of deception: A meta-analytic synthesis. *Applied Cognitive Psychology*, *20*, 421–446. doi:10.1002/acp.1190

Strömwall, L. A., Hartwig, P. A., & Granhag, P. A. (2006). To act truthfully: Nonverbal behaviour and strategies during a police interrogation. *Psychology, Crime & Law*, *12*, 207–219. doi:10.1080/10683160512331331328

Strömwall, L., & Willen, R. M. (2011). Inside criminal minds: Offenders' strategies when lying. *Journal of Investigative Psychology and Offender Profiling*, *8*, 271–281. doi:10.1002/jip.148

Undeutsch, U. (1982). Statement reality analysis. In A. Trankell (Ed.), *Reconstructing the past: The role of psychologists in criminal trials* (pp. 27–56). Deventer: Kluwer Academic.

Vrij, A. (2005). Criteria-based content analysis: A qualitative review of the first 37 studies. *Psychology, Public Policy, and Law*, *11*(1), 3–41. doi:10.1037/1076-8971.11.1.3

Vrij, A. (2008). *Detecting lies and deceit: Pitfalls and opportunities* (2nd ed.). New York, NY: Wiley.

Vrij, A., Akehurst, L., Soukara, S. & Bull, R. (2004). Let me inform you how to tell a convincing story: CBCA and reality monitoring as a function of age, coaching and deception. *Canadian Journal of Behavioural Science*, *36*(2), 113–126. doi:10.1037/h0087222

Vrij, A., Granhag, P. A., Mann, S., & Leal, S. (2011). Outsmarting the liars: Toward a cognitive lie detection approach. *Current Directions in Psychological Science*, *20*(1), 28–32. doi:10.1177/0963721410391245

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

Vrij, A., & Mann, S. (2006). Criteria-based content analysis: An empirical test of its underlying processes. *Psychology, Crime & Law*, *12*, 337–349. doi:10.1080/10683160500129007

Vrij, A., Mann, S., Kristen, S., & Fisher, R. P. (2007). Cues to deception and ability to detect lies as a function of police interview styles. *Law and Human Behavior*, *31*, 499–518. doi:10.1007/s10979-006-9066-4

Walsh, D., Milne, R., & Bull, R. (in submission). One way or another? Criminal investigators' belief regarding the disclosure of evidence to suspects in interviews.