Partners in Crime:
How Liars in Collusion Betray Themselves[1]

PÄR ANDERS GRANHAG[2] AND LEIF A STRÖMWALL
Göteborg University
Göteborg, Sweden

ANNA-CARIN JONSSON
Lund University
Lund, Sweden

The paradigmatic task for participants in studies on deception is to assess veracity on the basis of a single statement. However, in applied contexts, lie catchers are often faced with multiple statements (reported by one or several suspects). To appreciate this mismatch, we conducted a study where each member of 10 truth-telling pairs and 10 lying pairs (reporting fabricated alibis) was interrogated twice about an alibi. As predicted, lying pair members were more consistent between themselves than were truth-telling pair members, and single liars and truth tellers were equally consistent over time. Furthermore, truth tellers made more commissions than did liars. Although in line with our repeat vs. reconstruct hypothesis, these findings contrast sharply with beliefs held by professional lie catchers and recommendations found in literature on deception detection. The results are translated into an applied psycholegal context.

To reflect on questions of truth and deceit is an everyday activity (DePaulo, Kashy, Kirkendohl, Wyer, & Epstein, 1996). For some professional groups (e.g., police officers, judges), these reflections are often formalized into explicit assessments of veracity. Research mapping deception-detection performance shows that both presumed experts and laypeople are poor at distinguishing truth tellers from liars (Vrij, 2000). Furthermore, research has shown that believed indicators of deception are not in accordance with actual indicators (Strömwall & Granhag, 2003; Zuckerman, DePaulo, & Rosenthal, 1981). Hence, psychological research faces an important challenge in debunking false beliefs about how to detect

[2]Correspondence concerning this article should be addressed to Pär Anders Granhag, Department of Psychology, Göteborg University, Box 500, SE-405 30 Göteborg, Sweden. E-mail: pag@psy.gu.se

848

deceit, and in providing guidelines that will improve the quality of professional lie catchers' veracity assessments.

The body of psychological research conducted on interpersonal deception is voluminous (Vrij, 2000). However, the fact that lie catchers acting in real-life settings often must detect deceit on the basis of statements derived from several suspects—each having been interrogated repeatedly—has been overlooked by scholars within the field. Hence, research on deception has very little to offer when it comes to how, for example, stories fabricated by pairs in collusion differ from stories told by truth-telling pairs. Since it is far from rare that two (or more) persons are suspected of having committed a crime together and providing false alibis, such questions are of considerable psycholegal importance.

Research shows that perceptions of consistency mediate judgments of credibility (Greuel, 1992; Leippe, Manion, & Romanczyk, 1992). Furthermore, recent research shows that when lie catchers are asked to assess veracity on the basis of consecutive statements given by one and the same suspect, the most commonly reported cue by far is consistency (Granhag & Strömwall, 2000a, 2000b).

The main focus of the current study is to analyze the extent to which statements told by pairs reporting truthful accounts of an event differ from statements told by pairs reporting fabricated alibis. As will be further justified in the next section, we examine whether truthful and deceptive statements differ in terms of consistency. Our investigation covers both consistency within single suspects and consistency within pairs of suspects.

Consistency Between Statements

At least two lines of research must be taken into account as one deals with the question of whether truthful and deceptive accounts differ in terms of consistency. First, one must consider how repeated recall affects memory performance. The issue of repeatedly recalled stories received early attention in psychology through the work of Bartlett (1932), and several scholars have noted its relevance to forensic contexts (Penrod, Loftus, & Winkler, 1982). Summing up a large corpus of research, one of the most salient characteristics of memory, bearing on repeated recall, is its reconstructive nature (Tulving, 2000). Hence, for our context, truth tellers, who repeatedly try to retrieve a previously experienced event, can be expected to gain, lose, and change information over time (Baddeley, 1990; Loftus, 1979). Specifically, one effect of particular relevance is *reminiscence*, that is, when successive recall attempts produce previously unrecalled information (Payne, 1987). Turtle and Yuille (1994) have argued and shown that this effect is common in forensic contexts.

Second, in an attempt to establish a theoretical framework for deceptive behavior, three different approaches have been defined: the emotional (Ekman, 1985), the cognitive (Köhnken, 1989), and the attempted control approach

(DePaulo, 1992). These three approaches refer mainly to nonverbal deceptive behavior (Vrij, 1998). However, we have reason to believe that one of these approaches (the attempted control approach) bears on the type of verbal behavior relevant for the current study. The attempted control approach proposes that a liar will plan and control his or her behavior in order to appear as normal and credible as possible. The main prediction following this approach is that the liar will appear to be rehearsed (Vrij, 1998). Reasonably, this will hold for the nonverbal as well as for the verbal behavior expressed. This prediction fits nicely with the fact that a successful liar must have a good memory to remember what he or she has stated in previous interrogations (Granhag & Strömwall, 2000a) and that liars in collusion know that planning is crucial (Wagenaar & Dalderop, 1994).

Intertwining these two lines of research into one theoretically derived notion, the repeat versus reconstruct hypothesis has been suggested (Granhag & Strömwall, 1999, 2000a). This hypothesis rests on the assumption that liars will try to repeat what they have said in previous interrogations (or have agreed upon), while truth tellers will try to reconstruct what they at some point in time actually experienced. Reasonably, truth tellers, compared to liars, are less concerned with what they have said in previous interrogations. It can be argued that the repeat strategy used by liars will promote consistency and that the more recollective nature of truth tellers' memory work will, at least to some extent, undermine consistency.

Credibility Assessment and Verbal Behavior

The most widely used credibility assessment technique, focusing on verbal correlates of deception, is the criteria-based content analysis (CBCA). The CBCA, originally developed for analysis of children's reports, rests on a hypothesis stating that statements derived from memory of genuine experiences differ in content and quality from statements based on fantasy or invention (Undeutsch, 1989). In short, the CBCA is based on a number of different truth criteria (e.g., logical structure and quantity of details), and the assumption is that the more criteria are present, the higher the likelihood that the Statement is truthful. After a full CBCA has been conducted, it is suggested that a validity checklist should be used (Vrij, 2000). For the current context, one particular component of this checklist is of immediate interest: consistency with other statements. This component deals with inconsistencies between consecutive statements given by one witness, as well as inconsistencies between statements given by different witnesses. After reading the CBCA literature (e.g., Steller, 1989; Undeutsch, 1989), we found the description of this component to be vague (e.g., "Caution is required in order to demand not too much of persistence in a sequence of statements," Undeutsch, 1989, p. 115). Our contention is that CBCA advocates fail to make testable predictions in terms of consistency between statements.

Greuel, Brietzke, and Stadler (1999) stressed that the CBCA literature often neglects questions dealing with consistency. Nevertheless, CBCA scholars like Undeutsch and Greuel suggest that we should expect truthful statements to be consistent in terms of the salient aspects of the event (e.g., central persons and actions). This assumption is supported by Wagenaar, van Koppen, and Crombag (1993), who suggested that one criterion characterizing truthful consecutive statements is consistency in the relevant parts. In addition, a recent survey by Strömwall and Granhag (2003) shows that police officers, judges, and prosecutors subscribe to the belief that truthful statements are more consistent over time than are deceptive statements. A critical question is to what extent these assumptions are supported by empirical data. Much to our surprise, an inventory of psychological research on deception shows that the question of whether truthful and deceptive statements actually differ in terms of consistency has been given very little attention. In all, we found only two studies addressing the issue.

First, Wagenaar and Dalderop (1994), investigating consistency within pairs of suspects, sent six pairs to a zoo. Six other pairs were instructed to invent a story about a joint visit to the zoo. Later, all participants individually gave a free recall and answered a number of specific questions. Consistency within pairs was scored both in terms of explicitly stated facts (e.g., both members in the pair said "It rained in the morning") and facts that were mentioned in a more implicit manner (e.g., one pair member said "It rained in the morning and in the afternoon," while the other member said "In the afternoon it started to rain again"). Irrespective of type of scoring procedure used, the results showed that the lying pairs were significantly more consistent than were the truth-telling pairs.

Second, Granhag and Strömwall (2001), investigating consistency within a single suspect's consecutive statements, showed a staged robbery to 24 witnesses. Half of the witnesses were instructed to tell the truth, while the other half were told to distort the event in such a way that the actual victim was made to be the perpetrator. Each witness was then interrogated on three different occasions (3 hr, 4 days, and 11 days after the event). Analyzing different objective indicators of consistency (repeated details, omissions, and commissions), it was concluded that truthful and deceptive statements were equally consistent over time (Granhag & Strömwall, 2002). Note that the suggested repeat versus reconstruct hypothesis can account for the results obtained by Wagenaar and Dalderop (1994) and by Granhag and Strömwall (2002). That is, the repeat versus reconstruct hypothesis pertains to both single suspects interrogated repeatedly and to pairs of suspects.

The Present Study

Twenty pairs of friends volunteered to participate in an experiment on decision making. Unaware that they would later face in-depth questioning about the event, 10 pairs started off the experiment by having lunch at a restaurant. After returning

to the laboratory, each pair first completed a filler task together. Next, each pair was split and each pair member was instructed that he or she would now be interrogated about the lunch. The other 10 pairs did not get a free lunch. Instead, they were told to invent that they had had lunch together on a certain day. That is, they were instructed to fabricate an alibi that would withstand later interrogation. In sum, we had both lying and truth-telling pairs, and each pair member (hereafter referred to as *suspect*) was interrogated separately on two different occasions, with an interval of 7 days. By using this design, we made it possible to investigate consistency within single suspects, as well as consistency within pairs of suspects.

We used two different approaches in order to try to capture differences in terms of consistency. First, we conducted a textual analysis mapping the following three types of inconsistencies: contradictions (e.g., a car was first reported as red and then as blue), omissions (e.g., a knife was mentioned during the first interrogation, but not during the second), and commissions (a knife was mentioned during the second interrogation, but was left out during the first). Second, we collected ratings of consistency from participants who had watched the interrogations on video. We believe that such subjective ratings are of importance since a lie catcher's perception of the overall degree of consistency can instigate a miscarriage of justice. For example, a set of statements perceived as inconsistent can, in an unwarranted manner, evoke suspicion and lead to suspect-driven investigations (van Koppen, 1994). Similarly, a set of consecutive statements perceived as high in consistency can, in an equally unwarranted manner, subdue suspicion.

> *Hypothesis 1.* In view of our theoretically based repeat versus reconstruct hypothesis and previous findings (Wagenaar & Dalderop, 1994), we expect that both objective and subjective measures of consistency will show that lying pairs are at least as consistent as are truth-telling pairs.

> *Hypothesis 2.* In view of the same hypothesis and previous findings by Granhag and Strömwall (2002), we expect that single liars will be at least as consistent over time as are truth tellers.

Method

Participants

Videotaped statements from 40 psychology undergraduates (26 women, 14 men) were used. The students participated on a voluntary basis and were paid 150 SEK (approximately US$17) for their contribution. They were instructed to arrive at the laboratory in pairs for an experiment on decision making. Each pair came at a different time and was randomly assigned to be either a truthful pair or a lying pair of suspects.

Procedure

Truth-telling suspects. Upon their arrival, each truth-telling pair was given bogus information that the experiment was running late. Therefore, they were offered a free lunch at any of three nearby restaurants and were asked to come back within 1 hr. On returning, each pair was given a filler task, lasting approximately 15 min. Then, each pair was split up, and each pair member was informed of the real purpose of the experiment: that they were going to be interrogated about the lunch. They were instructed to answer the interrogator's questions as truthfully as possible.

Lying suspects. The so-called lying suspects also arrived at the laboratory in pairs. Each pair was seated in a room and received instructions stating that a crime had been committed and that they were suspected of committing it. They were told to state that they had been having lunch together at the time that the crime had been committed. In short, they were instructed to fabricate an alibi that would withstand later interrogation. Each pair was given 30 min to invent their story. They were then separated, performed the same filler task as the truth-telling suspects, and were interrogated. Following the interrogations, both lying and truth-telling suspects were told that they would be interrogated again 1 week later.

Interrogations. Interrogations were conducted by 2 persons, hired by the experimenters and blind to the veracity of the suspects. The first interrogation session consisted of a free-recall phase followed by 11 specific questions (Appendix). The questions were the same for both lying and truth-telling suspects. Each interrogation was highly structured and lasted approximately 5 min. All pairs returned for the second interrogation session 1 week later. The same interrogator as before then questioned each suspect. Each interrogation consisted of a free-recall phase followed by six specific questions (Appendix) and lasted approximately 4 min. All interrogations were videotaped. After each interrogation, the suspects were asked to rate on a 10-point scale ranging from 1 (*very low*) to 10 (*very high*) the degree of truth in the statement they had just given. After the second interrogation, the participants were debriefed and paid.

In this fashion, videotaped interrogations were conducted with 10 lying pairs of suspects and 10 truth-telling pairs of suspects, all interrogated twice. The videos were edited in order to fit with the presentation order in each of the viewing conditions (see the Ratings of Consistency section). All videotaped material was transcribed verbatim for further analyses.

Breaking Down the Text

To be able to examine possible differences between truthful and deceptive statements, we broke down the free-recall parts of the texts into information units and themes.

Information units. The first step was to remove half-spoken words and "ahs" and "mmms" from the statements. Next, we divided each specific statement into information units without altering the order of the narrative. We defined an *information unit* as the smallest possible segment of text giving a meaningful contribution to the narrative. An example of dividing the text into information units is: /We entered the restaurant/walked up to the counter/looked at the menu/. As the first example shows, the segments "We entered" and "the restaurant" were not considered to be giving a meaningful contribution on their own, but combined they were identified as making up one information unit. One person coded all statements and another person coded 20% of the statements. The proportion of interrater agreement for this particular step was .97.

Themes. The second step was to identify themes. A *theme*, in this context, is defined as a delimited topic made up of a number of information units (e.g., events taking place during the lunch or descriptions of persons and places). The following example shows theme building as seven information units were combined into the theme of "the waiter": /We had this really nice waiter/looked a bit Italian/quite cute/with brown curly hair/and brown eyes/and very cheerful/we really noticed him/.

For each suspect, the themes in Statement 1 were named and counted. Then, Statement 2 was checked for repeated themes, omissions (themes mentioned in Statement 1 but not in Statement 2), and commissions (themes not mentioned in Statement 1 but added in Statement 2).

With very few exceptions, a theme was made up of more than one information unit. For information units to be included in a theme, they had to concern one and the same event or description. One information unit could never be included in two different themes. A few information units had to be excluded since they were of a transitional nature (e.g., "and then . . .").

One person coded all of the statements, and another person coded 20% of the statements. Interrater agreement between two coders was 78%. That is, 78% of all information units were classified into the same themes by both coders. Another way of measuring the interrater agreement is to correlate the number of themes that both coders found in the statements; this analysis showed a correlation of .89.

Scoring of Contradictions

The full statements (free recall and specific questions) were analyzed for contradictions. We traced contradictions between (a) the two pair members' first statements, (b) the two pair members' second statements, and (c) each pair member's two statements. To be labeled a *contradiction*, the information had to be explicitly contradictory; for example, one person said, "I [Person A] ate fish and mashed potatoes," while the other person in the pair claimed "My friend [Person A] had chicken for lunch." Hence, neither omissions nor commissions were

included in this analysis. One person coded all of the statements, and another person coded 20% of the statements. For these statements, both coders found the same two contradictions. Overall, very few contradictions were found.

Ratings of Consistency

We collected ratings of consistency from 120 persons (48 men, 72 women). All were undergraduate students at Göteborg University and were paid SEK 50 (US$6) for their participation.

The raters watched the videotapes in small groups. They were instructed to focus on the consistency of the statements. The ratings were made on a 10-point scale ranging from 1 (*very low degree*) to 10 (*very high degree*). The 120 raters were randomly assigned to one of three viewing conditions, 40 in each. In the within single suspects condition, the raters were shown the two interrogations with one pair member, and then rated the consistency of that suspect's statement. In the within pairs of suspects–Statement 1 condition, raters were shown the first interrogations with both pair members, and rated the consistency. Finally, the raters in the within pairs of suspects–Statement 2 condition were shown the second interrogations with both pair members, and rated the consistency of those statements. Each observer watched either one pair of suspects or both interrogations with one suspect.

Results

Manipulation Check

To check whether the suspects complied with the instructions to tell the truth or to lie, a 2×2 (Statement Veracity: Truthful vs. Deceptive $\times$ Time: Statement 1 vs. Statement 2) ANOVA with repeated measurements on the last factor was conducted. A significant main effect for statement veracity was obtained, $F(1, 38) = 101.78, p < .001, \eta^2 = .73$. In both interrogation sessions, truth tellers rated the degree of truth in their statements significantly higher than did the liars. For Statement 1, the mean rated degree of truth was 9.80 (truth tellers) and 4.60 (liars); and for Statement 2, the corresponding means were 9.10 and 5.10, respectively. No significant effects were obtained for either the time factor or the interaction term.[3]

Consistency

First, we present four analyses of the free-recall parts of the transcribed statements, all pertaining to single suspects: (a) number of information units; (b) number of themes; (c) number of information units per theme; and (d) an analysis mapping repeated themes, omissions, commissions, and net and gross

---

[3]Regarding the number of words in the statements, no significant differences between the truthful and deceptive statements were found for either Statement 1 or Statement 2, both $ps > .30$.

Table 1

Mean Number of Themes, Information Units, and Information Units Per Theme by Truthful and Deceptive Statements

Level of analysis	Truthful		Deceptive	
	M	*SD*	*M*	*SD*
Number of information units				
Statement 1	38.60	33.45	44.50	39.45
Statement 2	33.85	17.75	29.80	29.76
Number of themes				
Statement 1	9.10	4.68	8.55	5.86
Statement 2	10.15	4.46	7.40	5.68
Number of information units/theme				
Statement 1	3.73	1.41	4.80	1.85
Statement 2	3.39	1.10	3.73	0.96

Note. The table contains analyses of the free-recall phase only.

recall. Next, we present an analysis focusing on pairs of suspects; namely, the number of themes mentioned by both pair members. Then, we present an analysis regarding contradictions in the suspects' statements. Finally, we report the results from the raters' perceived consistency. The main question throughout all analyses was as follows: Is there a difference, in terms of consistency, between truthful and deceptive statements?

Information units. A 2×2 (Statement Veracity: Truthful vs. Deceptive $\times$ Time: Statement 1 vs. Statement 2) mixed ANOVA was conducted to examine the number of information units in the statements. A significant main effect for the Time factor emerged, $F(1, 38) = 6.84$, $p < .05$, $\eta^2 = .15$, showing that significantly fewer information units were found in Statement 2 than in Statement 1. The Statement Veracity factor and the Statement Veracity $\times$ Time interaction were not significant. Table 1 presents the descriptive statistics.

Themes. In order to find out if there were any differences in the number of themes reported, a 2×2 (Statement Veracity: Truthful vs. Deceptive $\times$ Time: Statement 1 vs. Statement 2) mixed ANOVA was conducted. No main effects were significant, but the Statement Veracity $\times$ Time interaction was significant, $F(1, 38) = 4.00$, $p = .05$, $\eta^2 = .10$. This interaction indicates that truthful statements contained more themes in Statement 2 compared to Statement 1, and that deceptive statements contained fewer themes in Statement 2 than in Statement 1. Table 1 shows means and standard deviations.

Table 2

Means and Standard Deviations for Features of the Theme Analysis by Truthful/Deceptive Statements

Analysis feature	Truthful		Deceptive		
	M	SD	M	SD	t
Total number of themes					
Statement 1	9.10	4.68	8.55	5.86	0.33
Statement 2	10.15	4.46	7.40	5.68	1.70
Repeated themes	6.35	3.48	5.35	4.69	0.76
Omissions	2.75	2.40	3.20	2.38	0.60
Commissions	3.80	2.55	2.05	1.67	2.57*
Net recall	1.05	4.20	-1.15	2.56	2.00*
Gross recall	12.90	4.93	10.60	6.82	1.22

*$p < .05$.

Information units per theme. The number of information units making up each theme was averaged for each suspect. A 2 × 2 (Statement Veracity: Truthful vs. Deceptive × Time: Statement 1 vs. Statement 2) mixed ANOVA was conducted, and it showed a significant main effect of time, $F(1, 38) = 11.46, p < .01, \eta^2 = .23$. The themes in Statement 1 contained more information units than did the themes in Statement 2 (see Table 1 for descriptives). The Statement Veracity factor was quite close to significance, $F(1, 38) = 3.41, p = .08, \eta^2 = .08$, with liars reporting a larger number of information units per theme. Finally, the interaction term was also quite near a significant effect, $F(1, 38) = 3.09, p = .09, \eta^2 = .08$. The difference between truth tellers and liars at Statement 1 was substantially reduced in Statement 2.

For each suspect, the themes in Statement 1 were named and counted. In order to test our suggested repeat versus reconstruct hypothesis, Statement 2 was checked for repeated themes, omissions, and commissions. Furthermore, we calculated the net and gross recall in terms of number of themes. Mean values, standard deviations, and *t* values are presented in Table 2.

Repeated themes, omissions, and commissions. For the repeated themes, the *t* test did not show a significant difference between truth tellers' ($M = 6.35$) and liars' ($M = 5.35$) statements. Likewise for the omissions, no difference was found between truth tellers' ($M = 2.75$) and liars' ($M = 3.20$) statements. For the commissions, the means were significantly separated: Truth-telling suspects ($M = 3.80$) included more commissions than did lying suspects ($M = 2.05$), $t(38) = 2.57, p < .05$.

Table 3

Percentages of Overlapping Themes for Statements 1 and 2 and by Truthful/Deceptive Pairs

	Truthful pairs		Deceptive pairs		
	M	*SD*	*M*	*SD*	*t*
Statement 1	38.17	15.57	60.58	18.06	2.97**
Statement 2	41.65	18.01	55.34	12.35	1.98

Note. The table entries refer to the percentage of overlapping themes that both pair members mentioned in their statements.
**$p < .01$.

Net and gross recall. We also compared truthful and deceptive statements in terms of net recall (for our context, the number of commissions minus the number of omissions), and gross recall (for our context, the number of themes mentioned in Statement 1 plus the number of commissions). Truthful ($M = 1.05$) and deceptive ($M = -1.15$) statements differed significantly in terms of net recall, $t(38) = 2.00$, $p = .05$. Notably, the net recall was positive for truthful statements, and it was negative for deceptive statements. Furthermore, the gross recall was nonsignificantly higher for truthful statements ($M = 12.90$) than for deceptive statements ($M = 10.60$).

Number of themes overlapping between pair members. Note that the analyses presented so far concern single suspects, not pairs. We compared truth-telling and lying pairs with respect to the extent to which the two pair members' statements overlapped. The number of overlapping themes was put in relation to all themes mentioned by the two pair members (i.e., the percentage of overlapping themes). A 2 × 2 (Statement Veracity: Truthful vs. Deceptive Pairs × Overlapping Themes: Statement 1 vs. Statement 2) mixed ANOVA was conducted. The Statement Veracity factor was significant, $F(1, 18) = 12.73$, $p < .01$, $\eta^2 = .41$, showing that the number of overlapping themes was significantly higher for lying pairs than for truth-telling pairs. The Overlapping Themes factor and the interaction term were not significant ($ps > .40$). Means and standard deviations are presented in Table 3. Tests for each statement showed that the overall significant result could be attributed primarily to Statement 1, with deceptive statements showing significantly more overlapping themes than truthful statements, paired samples $t(18) = 2.97$, $p < .01$. For Statement 2, the difference was in the same direction, and the t test just bordered on significance, paired samples $t(18) = 1.98$, $p = .06$.

Contradictions. We traced contradictions both within single suspects and between pair members. Within single suspects, only 3 contradictions were found;

Table 4

Mean Values of Consistency for Each Viewing Condition and by Truthful/Deceptive Statements

Viewing condition	Truthful		Deceptive		Univariate *t* values
	M	*SD*	*M*	*SD*	
Within single suspects	7.75	1.92	8.65	0.99	1.87
Within pairs of suspects					
Statement 1	7.50	1.47	8.70	0.73	3.27**
Statement 2	7.55	1.50	8.05	1.19	1.17

Note. The ratings of consistency were made on a 10-point scale ranging from 1 (*low*) to 10 (*high*).
**$p < .01$.

and between pair members, 14 contradictions were identified. Out of these 17 contradictions, liars committed 9 and truth tellers committed 8. As the total number of contradictions was very low, we did not analyze these further.

Perceived consistency. To further examine the consistency of suspects' statements, we collected data from participants who rated the content of the video recordings. Their perceived consistency was analyzed with a 2 × 3 (Statement Veracity: Truthful vs. Deceptive × Viewing Condition: Within Single Suspects vs. Within Pairs of Suspects–Statement 1 vs. Within Pairs of Suspects–Statement 2) between-subjects ANOVA. The Viewing Condition factor and the Statement Veracity × Viewing Condition interaction were not significant. The Statement Veracity factor showed a significant effect, $F(1, 114) = 12.27, p < .001, \eta^2 = .10$. The means (Table 4) indicate that the deceptive statements were perceived as more consistent than were the truthful statements. Further tests (univariate Bonferroni corrected *t* tests; see the last column of Table 4) reveal that this difference was most prevalent for those who watched the first two interrogations with both pair members (i.e., the viewing condition within pairs of suspects–Statement 1), $t(38) = 3.27, p < .01$.

Discussion

In applied settings, lie catchers often assess veracity on the basis of multiple statements, both in the form of consecutive statements from single suspects and statements produced by several persons suspected of committing a crime (Gudjonsson, 1992; Shuy, 1998). The full importance of this fact has not been appreciated within the deception-detection paradigm.

The current study aimed to map the extent to which statements given by pairs reporting a false alibi differed from statements produced by pairs giving a truthful account of an event that they had experienced together. Critically, all truth tellers and liars were interrogated individually on two separate occasions. Specifically, we investigated whether truthful and deceptive statements differed in terms of consistency within single suspects and consistency within pairs of suspects.

The lying pairs were asked to fabricate an alibi involving them having lunch together. Obviously, a lunch scenario is a strongly scripted event (Schank & Abelson, 1977). Hence, the liars' task of fabricating a credible alibi probably was facilitated by the utilization of such a readymade script. However, and importantly, we believe that the use of strongly structured events in inventing false alibis is common in real-life situations. In short, placing oneself in a familiar environment, performing a highly scripted activity is probably a very wise route to choose when fabricating an alibi.

Consistency Between Statements

Investigating consistency between statements is a rather difficult task (for a similar view, see Wagenaar & Dalderop, 1994). In short, there is no principled index mirroring the degree of consistency between two (or more) statements. In the current study, we used two different approaches to examine differences in terms of consistency between truthful and deceptive statements. First, and in order to obtain an objective measure, we performed a textual analysis of themes (mapping number of repeated themes, omissions, commissions, and contradictions). In a study by Ghetti, Eisen, Goodman, Qin, and Davis (1998) that investigated consistency in children's allegations of sexual and physical abuse, the authors searched for consistencies or inconsistencies for a fixed number of prespecified content categories (e.g., actions attributed to the perpetrator). In contrast, our textual analysis included every single theme that was identified in the statements. In short, compared to Ghetti et al., we performed a more data-driven type of consistency analysis. Second, to obtain a complementary index of consistency, we asked participants first to watch the interrogations on video and then to rate the degree of consistency. In the introduction, we argued why such a subjective measure is important in applied psycholegal contexts.

Furthermore, it should be noted that inconsistencies can be studied in many different contexts. The current study dealt with the extent to which lying and truth-telling pair members are consistent between themselves, and to which liars' and truth-tellers' consecutive statements are consistent over time. Hence, the present study did not focus on, for example, inconsistencies found within one suspect's single statement or discrepancies between a suspect's verbal and nonverbal behavior.

Consistency within single suspects. The subjective ratings that were collected indicate that truthful and deceptive statements were perceived as equally consistent over time. Corroborating this finding, our textual analysis shows that there were no significant differences between truth tellers and liars in terms of number of repeated themes, omissions, and contradictions. It was found that truth tellers added significantly more new themes to their initial statements than did liars. That is, we found a stronger reminiscence effect for truthful statements than for deceptive statements. Furthermore, the average net recall was significantly higher for truthful statements than for deceptive statements. These two findings support the idea that asking a person to elaborate on some previously reported information is less likely to result in the addition of information if the story is fabricated than if the story is truthful (Jackson & Granhag, 2001; Soppe, 1995, as cited in Vrij, 2000). Recently, Shuy (1998), without any further qualifications, claimed that a memory improving over time belongs to a liar. Empirical data suggest that Shuy's statement is wrong and that the opposite is more correct.

In sum, the prediction that single liars would produce statements that were at least as consistent as truth tellers was confirmed by both the subjective ratings and the textual analysis. In addition, our textual analysis shows that truthful and deceptive statements did not differ in terms of number of themes, information units, and information units per theme. However, truth tellers showed an increase in the number of themes reported over time, whereas liars showed a decrease in the number of themes reported, as indicated by a significant interaction. Both liars and truth tellers showed a decrease in reported information units over time.

Consistency within pairs of suspects. We predicted that lying pair members would be at least as consistent between themselves as truth-telling pair members. Indeed, the textual analysis shows that the relative number of overlapping themes was higher for lying pairs than for truth-telling pairs (this difference was significant for statements produced during the first interrogation and was very close to significant for statements produced during the second interrogation). Fully in line with this pattern, the lying pairs were rated as more consistent than truth-telling pairs (significantly for statements produced during the first interrogation, but not significantly for statements produced during the second interrogation). Hence, both the objective and subjective measures of consistency supported our prediction.

Furthermore, the analysis shows that both lying and truth-telling pair members contradicted themselves to a low extent. In order to investigate why, we scored the number of times the suspects answered, "I don't know" or "I can't remember." We found that both truth tellers (27) and liars (24) seldom used these loopholes. (In total, 680 specific questions were asked to the single suspects.) Speculatively, this finding is a result of the lying pairs' careful planning,

and their strategy to place themselves in a familiar environment, doing an everyday activity. Hence, the liars enabled themselves to answer the specific questions without resorting to more "I don't know" answers or contradictions than the truth tellers.

The repeat versus reconstruct hypothesis. The repeat versus reconstruct hypothesis outlined in the introduction suggests that liars will try to repeat what they stated during previous interrogations, whereas truth tellers will try to reconstruct what they actually experienced. With reference to this hypothesis, deceptive statements, compared to truthful ones, can be expected to be equally consistent (Granhag & Strömwall, 1999). On a general level, the results of the current study lend strong support to the repeat versus reconstruct hypothesis. Specifically, the results suggest that (a) the repeat component of the hypothesis might be more useful in distinguishing lying pairs from truth-telling pairs than in distinguishing single liars from single truth tellers; and (b) the reconstructive component of the hypothesis manifests itself mainly in terms of commissions.

To further evaluate the repeat versus reconstruct hypothesis, the effects of both time and information density must be examined. With reference to basic memory principles such as the forgetting curve (Baddeley, 1990), it can be speculated that the hypothesis is even more valid if there is a long retention interval between the to-be-remembered event and the first interrogation, or if there are long time lapses between a suspect's report of a very detailed initial statement and the follow-up interrogation. Furthermore, in future research, it might be of interest to examine possible differences between situations where the suspects are forewarned that they are be interrogated repeatedly, as in the present study, and situations where they are not warned.

In the current study, and fully in line with previous research (Granhag & Strömwall, 2002; Wagenaar & Dalderop, 1994), we found larger differences in terms of consistency when comparing truth-telling and lying pairs of suspects than when comparing truth-telling and lying single suspects. One possible explanation for this result is that the lying pairs were given time to fabricate their alibi before being interrogated for the first time. The truth-telling pairs, who actually had had lunch together, were not given any opportunity to discuss and recapitulate what really happened during the lunch. In fairness, it might very well be that the pattern of results would have been different if we had used a design where the truth tellers had been given the chance to collude or the liars were not given the chance to collude. However, and importantly, we would like to raise two arguments in defense of the design employed in the current study. First, partners in crime know that it is essential to agree on what to say in the interrogations to follow. Hence, they will make sure to have an opportunity to collude. Second, truth-telling suspects normally do not feel the same need to reach an agreement between them before being interrogated. Therefore, we argue that the design of the current study is ecologically valid.

Psycholegal Implications

Research shows that both professional lie catchers, such as police officers (Greuel, 1992) and prosecutors and judges (Strömwall & Granhag, 2003), agree on the heuristic stating that inconsistency implies deception and consistency implies truth. In sharp contrast to this belief, the results of the current study indicate that (a) statements produced by liars in collusion are more consistent than are statements produced by truth tellers who have experienced an event together, and (b) consecutive statements given by single liars and single truth tellers are equally consistent over time. These findings support our repeat versus reconstruct hypothesis, as well as previously reported results (Granhag & Strömwall, 2000a; Wagenaar & Dalderop, 1994).

Findings reported by Strömwall, Granhag, and Jonsson (in press) show that lie catchers assessing the veracity of pairs of suspects—where both pair members had been interrogated repeatedly—primarily attend to whether the two suspects are consistent between themselves, compared to whether each of the two suspects is consistent within himself or herself. Critically, the absolute majority of these lie catchers seem to hold the belief that truth-telling pair members are more consistent between themselves than are lying pair members. Our finding that truth-telling pair members were less consistent between themselves strongly questions the validity of this belief.

An important psycholegal question is whether or not it is possible to identify systematic differences between truthful and deceptive statements in terms of inconsistencies. The results of the current study suggest that the reply to this question might be positive. That is, we found that truth tellers added significantly more new information to their initial statements than did liars; that is, we found a stronger reminiscence effect for truthful statements. Importantly, this result corroborates previous findings (Jackson & Granhag, 2001; Soppe, 1995, as cited in Vrij, 2000). Obviously, more research is needed in order to resolve this question.

A common guideline for assessing veracity on the basis of consecutive statements, also supported by the CBCA framework, is that most truthful reports are consistent when it comes to the salient aspects of the statement, such as central persons and actions (Greuel et al. 1999; Shuy, 1998; Undeutsch, 1989; Wagenaar et al., 1993). Since our textual analysis shows that the same characteristic also seems to hold for deceptive statements, it might be that this recommendation does not carry much weight. A recently explored alternative to the CBCA technique is the reality monitoring (RM) approach, originally developed by Johnson and colleagues (e.g., Johnson & Raye, 1981). In brief, the assumption that memories for real events (in contrast to memories for imagined events) are rich in cues that are linked to their perceptual origin (e.g., colors, sensory properties) and contextual information (e.g., spatial, temporal) has received strong empirical support (Johnson, Foley, Suengas, & Raye, 1988; Johnson & Raye, 1981). Lately, several

researchers have suggested that the RM framework could be used as a tool for lie detection (e.g., Hernandez-Fernaud & Alonzo-Quecuty, 1997; Sporer, 1997), and experimental and field studies show a promising pattern of results (for a recent summary, see Vrij, 2000). In addition, the RM criteria are fewer and easier to teach and learn than are the CBCA criteria (Vrij, 2000). However, as was pointed out by Davies (2001), considerable work is required before the validity of RM as a lie-detection technique can be determined.

Translating our results into an applied psycholegal context, three basic recommendations can be offered. First, when a set of consecutive statements is found to be consistent, it should not be credited automatically. Instead, it should be kept in mind that liars often try very hard to remember what they have said in previous interrogations. Second, multiple suspects (or witnesses) sometimes report statements that, between them, reach an extraordinarily high level of consistency. This could indicate that the suspects are in collusion and have agreed on what to report. Obviously, in itself, this is not a waterproof sign of deception, but should lead to further investigation. Third, inconsistencies between two (or more) statements should not be interpreted immediately as incriminating. Reasonably, truthful statements are characterized by a natural inconsistency; successive recall attempts often result in new information being added to previously given information (reminiscence). If two suspects clearly contradict each other (or if a single suspect contradicts himself or herself), naturally this should evoke suspicion. It should be noted that because of the paucity of research on this particular issue of deception, these recommendations are based on a relatively meager number of empirical studies.

In conclusion, the combined evidence suggests that the consistency heuristic as used by most lie catchers is not very helpful in discerning truthful and deceptive statements (Strömwall et al., in press). However, we do not recommend that triers of facts should simply ignore issues of consistency. Such a recommendation would be too hasty and probably very ineffective. Instead, we hope that future research on this issue will result in empirically based guidelines useful for lie catchers assessing veracity on the basis of multiple statements. However, until then, we suggest that lie catchers exercise the consistency heuristic with great caution.

References

Baddeley, A. (1990). *Human memory: Theory and practice*. Hove, UK: Lawrence Erlbaum.

Bartlett, F. C. (1932). *Remembering: A study in experimental and social psychology*. Cambridge, UK: Cambridge University Press.

Davies, G. M. (2001). Is it possible to discriminate true from false memories? In G. M. Davies & T. Dalgleish (Eds.), *Recovered memories: Seeking the middle ground* (pp. 153-174). Chichester, UK: John Wiley & Sons.

DePaulo, B. M. (1992). Nonverbal behavior and self-presentation. *Psychological Bulletin, 111*, 203-243.

DePaulo, B. M., Kashy, D. A., Kirkendohl, S. E., Wyer, M. M., & Epstein, J. A. (1996). Lying in everyday life. *Journal of Personality and Social Psychology, 70*, 979-995.

Ekman, P. (1985). *Telling lies: Clues to deceit in the marketplace, politics, and marriage.* New York, NY: W. W. Norton.

Ghetti, S., Eisen, M. L., Goodman, G., Qin, J., & Davis, S. L. (1998, August). *Consistencies and inconsistencies in children's allegations of sexual and physical abuse.* Paper presented at the 24th International Congress of Applied Psychology, San Francisco, CA.

Granhag, P. A., & Strömwall, L. A. (1999). Repeated interrogations: Stretching the deception detection paradigm. *Expert Evidence, 7*, 163-174.

Granhag, P. A., & Strömwall, L. A. (2000a). Deception detection: Examining the consistency heuristic. In C. M. Breur, M. M. Kommer, J. F. Nijboer, & J. M. Reintjes (Eds.), *New trends in criminal investigation and evidence II* (pp. 309-321). Antwerp, Belgium: Intersentia.

Granhag, P. A., & Strömwall, L. A. (2000b). Effects of preconceptions on deception detection and new answers to why lie-catchers often fail. *Psychology, Crime, and Law, 6*, 197-218.

Granhag, P. A., & Strömwall, L. A. (2001). Deception detection based on repeated interrogations. *Legal and Criminological Psychology, 6*, 85-101.

Granhag, P. A., & Strömwall, L. A. (2002). Repeated interrogations: Verbal and nonverbal cues to deception. *Applied Cognitive Psychology, 16*, 243-257.

Greuel, L. (1992). Police officers' beliefs about cues associated with deception in rape cases. In F. Lösel, D. Bender, & T. Bliesener (Eds.), *Psychology and law: International perspectives* (pp. 234-239). Berlin, Germany: Walter de Gruyter.

Greuel, L., Brietzke, S., & Stadler, M. A. (1999, July). *Credibility assessment: New research perspectives.* Paper presented at the American Psychology-Law Society/European Association of Psychology and Law Conference on Psychology and Law, Dublin, Ireland.

Gudjonsson, G. H. (1992). *The psychology of interrogations, confessions, and testimony.* Chichester, UK: John Wiley & Sons.

Hernandez-Fernaud, E., & Alonzo-Quecuty, M. L. (1997). The cognitive interview and lie detection: A new magnifying glass for Sherlock Holmes? *Applied Cognitive Psychology, 11*, 55-68.

Jackson, J. L., & Granhag, P. A. (2001). *Who went to the cinema? Children's ability to coach friends and deceive parents.* Manuscript in preparation, Netherlands Institute for the Study of Criminality and Law Enforcement, Leiden, The Netherlands.

Johnson, M. K., Foley, M. A., Suengas, A. G., & Raye, C. L. (1988). Phenomenal characteristics of memories for perceived and imagined autobiographical events. *Journal of Experimental Psychology: General, 117*, 371-376.

Johnson, M. K., & Raye, C. L. (1981). Reality monitoring. *Psychological Review, 88*, 67-85.

Köhnken, G. (1989). Behavioral correlates of statement credibility: Theories, paradigms, and results. In H. Wegener, F. Lösel, & J. Haisch (Eds.), *Criminal behavior and the justice system: Psychological perspectives* (pp. 271-289). New York, NY: Springer-Verlag.

Leippe, M. R., Manion, A. P., & Romanczyk, A. (1992). Eyewitness persuasion: How and how well do fact finders judge the accuracy of adults' and children's memory reports? *Journal of Personality and Social Psychology, 63*, 181-197.

Loftus, E. F. (1979). *Eyewitness testimony.* Cambridge, MA: Harvard University Press.

Payne, D. G. (1987). Hypermnesia and reminiscence in recall: A historical and empirical review. *Psychological Bulletin, 101*, 5-27.

Penrod, S., Loftus, E. F., & Winkler, J. (1982). The reliability of eyewitness testimony: A psychological perspective. In N. L. Kerr & R. M. Bray (Eds.), *The psychology of the courtroom* (pp. 119-168). San Diego, CA: Academic Press.

Schank, R. C., & Abelson, R. (1977). *Scripts, plans, goals, and understanding.* Hillsdale, NJ: Lawrence Erlbaum

Shuy, R. W. (1998). *The language of confession, interrogation, and deception.* Thousand Oaks, CA: Sage.

Sporer, S. L. (1997). The less traveled road to the truth: Verbal cues in deception detection in accounts of fabricated and experienced events. *Applied Cognitive Psychology, 11*, 373-397.

Steller, M. (1989). Recent developments in statement analysis. In J. C. Yuille (Ed.), *Credibility assessment* (pp. 135-154). Dordrecht, The Netherlands: Kluwer Academic Publishers.

Strömwall, L. A., & Granhag, P. A. (2003). How to detect deception? Arresting the beliefs of police officers, prosecutors, and judges. *Psychology, Crime, and Law, 9*, 19-36.

Strömwall, L. A., Granhag, P. A., & Jonsson, A.-C. (in press). Deception among pairs: "Let's say we had lunch and hope they will swallow it!" *Psychology, Crime, and Law.*

Tulving, E. (2000). Concepts of memory. In E. Tulving & F. I. M. Craik (Eds.), *The Oxford handbook of memory* (pp. 33-43). New York, NY: Oxford University Press.

Turtle, J. W., & Yuille, J. C. (1994). Lost but not forgotten: Repeated eyewitness recall leads to reminiscence but not hypermnesia. *Journal of Applied Psychology, 79*, 260-271.

Undeutsch, U. (1989). The development of statement reality analysis. In J. C. Yuille (Ed.), *Credibility assessment* (pp. 101-119). Dordrecht, The Netherlands: Kluwer Academic Publishers.

van Koppen, P. J. (1994). From police information to miscarriages of justice. In N. K. Clark & G. M. Stephenson (Eds.), *Issues in criminological and legal psychology. Vol. 21. Rights and risks: The application of forensic psychology* (pp. 11-20). Leicester, UK: The British Psychological Society for the Division of Criminological and Legal Psychology.

Vrij, A. (1998). Nonverbal communication and credibility. In A. Memon, A. Vrij, & R. Bull (Eds.), *Psychology and law: Truthfulness, accuracy, and credibility* (pp. 32-58). London, UK: McGraw-Hill.

Vrij, A. (2000). *Detecting lies and deceit.* Chichester, UK: John Wiley & Sons.

Wagenaar, W. A., & Dalderop, A. (1994). *Remembering the zoo: A comparison of true and false stories told by pairs of witnesses.* Unpublished manuscript, Department of Experimental Psychology, Leiden University, The Netherlands.

Wagenaar, W. A., van Koppen, P. J., & Crombag, H. F. M. (1993). *Anchored narratives. The psychology of criminal evidence.* New York, NY: St. Martin's Press.

Zuckerman, M., DePaulo, B. M., & Rosenthal, R. (1981). Verbal and nonverbal communication of deception. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 14, pp. 1-57). New York, NY: Academic Press.

Appendix

Questions in the Interrogations

Initially in each interrogation, the person conducting the interrogation asked the suspects to give a free-recall narrative of the lunch: "Tell me about this lunch in full detail, take your time" (translation from Swedish). Then, the interrogation continued with a set of specific questions.

Specific questions asked in the first interrogation:

1. Did you have to wait a long time for the food?
2. What did your friend have for lunch?
3. What did your friend drink?
4. How certain are you about that statement?
5. Did you order coffee after the meal?
6. Were there tablecloths on the tables?
7. Can you mention some of the topics the two of you discussed?
8. Was the restaurant crowded?
9. How certain are you about that statement?
10. Was smoking allowed where you were sitting?
11. Have you been to this particular restaurant before?

Specific questions asked in the second interrogation:

1. Do you remember what you were wearing on this particular day?
2. Do you remember what your friend was wearing?
3. Was the restaurant crowded?
4. How certain are you about that statement?
5. Did you have to wait a long time for the food?
6. Do you remember the appearance of anyone working at the restaurant?

Note. Liars and truth tellers were asked the same questions.