December 29, 2017 – Accepted for publication in *Law and Human Behavior*

Truth-tellers stand the test of time and contradict evidence less than liars, even months after a crime

Divya Sukumar,[1] Kimberley A. Wade,[1] and Jacqueline S. Hodgson,[2]

University of Warwick, United Kingdom

Author Note

1 Department of Psychology, University of Warwick, Coventry, CV4 7AL, UK.

2 School of Law, University of Warwick, Coventry, CV4 7AL, UK

Correspondence concerning this article should be addressed to Kimberley A. Wade, Department of Psychology, University of Warwick, Coventry, UK, CV4 7AL. Email: K.A.Wade@warwick.ac.uk

Word count: 6756

Abstract

When deceptive suspects are unaware of the evidence the police hold against them, they contradict that evidence more than truthful suspects do – a useful cue to deception. But given that, over time, truthful suspects might forget the past and also contradict the evidence, how effective are lie detection techniques that rely on such inconsistencies when suspects are questioned months after a crime? In Experiment 1, people committed a theft (liars) or a benign activity (truth-tellers) in a university bookshop. Shortly after or two months later, we questioned them about their bookshop visit without informing them of the evidence implicating them in the theft. Though truth-tellers contradicted some evidence after both time delays, liars always contradicted the evidence more than did truth-tellers. In Experiment 2, we presented the mock suspects' responses to an independent group of laypeople and asked them to rate how deceptive the suspects were. Laypeople rated liars as more deceptive than truth-tellers after both time delays, but also rated truth-tellers questioned two months after the crime as more deceptive than truth-tellers questioned shortly after the crime. These findings suggest that liars' tendency to distance themselves from a crime might outweigh any memory decay that truth-tellers experience in the two months following a crime. As a result, the extent of a suspect's contradictions with the evidence could still be diagnostic of deception even after an extended time delay.

Keywords: deception detection; strategic use of evidence; time delay.

Public significance statement: When liars and truthful suspects were questioned about a mock crime either immediately after or two months after the crime, liars contradicted the incriminatory evidence more often than did truth-tellers. Although truthful suspects may also be perceived as lying, these findings suggest that withholding evidence in a police interview and considering the degree to which a suspect contradicts the evidence could help with deception detection in interviews that take place months after a crime.

Truth-tellers stand the test of time and contradict evidence less than liars, even months after a crime

Do you remember what you were doing exactly two months ago? Maybe not. Now imagine you are suddenly a suspect in a criminal case and the police ask you for an alibi for that day. You are probably struggling to remember the details. You might even say something that contradicts the police's evidence. The obvious danger for truthful suspects, like yourself, is that forgetting the past and contradicting the evidence could make you look like a liar. This scenario is not as farfetched as it may seem: We know that police around the world are instructed to rely on inconsistencies between suspects' statements and the available evidence to detect if suspects are lying (e.g. Association of Chief Police Officers, 2014; Hartwig, Granhag, Stromwall, & Kronkvist, 2006; Luke et al., 2016). In this paper, we ask whether lie detection techniques that hinge on the consistency between a suspect's statements and the evidence the police hold may be effective after a long delay.

It is unsurprising that the police capitalize on verbal cues to detect deception: A growing body of psychological research suggests that, when unaware of the evidence, liars are more likely to make statements that contradict the evidence than are truth-tellers (e.g., Clemens et al., 2010; Clemens, Granhag, & Stromwall, 2011; Hartwig, Granhag, Stromwall, & Vrij, 2005). This difference between liars and truth-tellers can be attributed to the different counter-interrogation strategies that suspects adopt when being questioned about a crime (Granhag & Hartwig, 2008, 2015). For liars, incriminating evidence is a threat and the interviewer might or might not possess such evidence (Granhag & Hartwig, 2008; Hartwig, Granhag & Luke, 2014). To deal with this threat and to appear credible, liars either avoid mentioning or deny any connection to the crime during the interview. In doing so, liars can unknowingly contradict the evidence that links them to the crime, known as *statement-evidence inconsistencies* (Hartwig et al., 2006).

In contrast, truth-tellers are typically more consistent with the interviewer's evidence because they tend to be forthcoming with information and try to tell their story as it happened (Colwell, Hiscock-Anisman, Memon, Woods, & Michlik, 2006; Stromwall, Hartwig, & Granhag, 2006). Truth-tellers' forthcoming approach may reflect a belief that people get what they deserve, the *belief in a just world* phenomenon (Lerner, 1980), coupled with a tendency to assume that their inner states, thoughts, and emotions are evident to others, the *illusion of transparency* (Gilovich, Savitsky, & Medvec, 1998). Together, these concepts could explain why innocent suspects may come to believe that by talking to the police their innocence will shine through and justice will prevail (Kassin, 2005; Kassin & Norwick, 2004). For instance, in one experiment, a detective accused students, some innocent and some guilty, of stealing $100 and informed them of their right to silence. While only 36% of the guilty students waived their right to silence, a striking 81% of the innocent students waived their right to silence, chose to talk to the detective, and explained that they "did nothing wrong" and "didn't have anything to hide" (Kassin & Norwick, 2004, p. 216). Similarly, truth-tellers are likely to admit their presence at the crime scene even if the interviewer has not informed them of any evidence linking them to the crime (Hartwig et al., 2014). In this way, truth-tellers tend to make fewer statement-evidence inconsistencies than do liars.

Accordingly, psychology researchers have developed a new repertoire of police interviewing techniques that use suspects' inconsistencies as a means of detecting deception (e.g., Leins, Fisher, & Vrij, 2012; Leins, Fisher, Vrij, Leal, & Mann, 2011; Vrij et al., 2009). One such technique with growing empirical support is the Strategic Use of Evidence (SUE) approach in which interviewers ask suspects to freely recall their activity during the time of the crime and to answer specific questions (e.g., "Did you see a briefcase?"; "Did you handle a briefcase?") before disclosing to the suspect the evidence that implicates them in the crime (Hartwig et al., 2014). For instance, in one study, mock suspects either stole a wallet from a

briefcase in a bookshop and lied about it to the interviewer or visited the bookshop in search of a hole-punch in a box underneath the same briefcase and told the truth about it to the interviewer (Hartwig et al., 2006). The study was set up so that evidence, such as the mock suspects' fingerprints found on the briefcase containing the wallet, implicated both the deceptive and truthful suspects in the theft of the wallet. Interviewers trained in SUE disclosed this evidence only after they had questioned the mock suspects on their activity in the bookshop. In these interviews, deceptive mock suspects made more statement-evidence inconsistencies than truthful mock suspects, and as a result, trained interviewers accurately detected 85.0% of truthful mock suspects and 85.7% of deceptive mock suspects. Indeed, a meta-analysis of eight studies found a large difference (d = 1.89) between deceptive and truthful mock suspects' statement-evidence inconsistencies when evidence was disclosed late in the interview (Hartwig et al., 2014). These findings suggest that statement-evidence inconsistencies could be a robust and diagnostic cue to when suspects are lying.

Further research is needed, however, to explore whether lie detection techniques that rely on suspects' inconsistencies, such as SUE, might work in a variety of forensically relevant conditions – including when suspects are questioned weeks or months after a crime (e.g., Shawyer, Milne, & Bull, 2009; see also Sukumar, Hodgson, & Wade, 2016 and Sukumar, Wade, & Hodgson, 2016 for other concerns about the strategic disclosure of evidence in practice). Up until now, the time delays used in published strategic evidence disclosure studies have been typically short. Upon reviewing the literature, we found 22 published studies in which mock suspects were questioned with some variant of the SUE technique (see Appendix for details). In 20 of these studies, suspects were questioned within one hour of the activity that they needed to lie or tell the truth about. The only exceptions were Hartwig et al. (2005) with a one week delay and McDougall and Bull (2015) with a delay of 7–10 days. Yet, over a longer delay, say, several weeks or months, truthful suspects

might forget what they were doing and also contradict the evidence, leading them to be mistaken for deceptive suspects.

Indeed, truthful suspects' ability to respond consistently with the evidence is likely to be compromised over time given that information encoded in memory can be rapidly forgotten, and over time, becomes increasingly difficult to retrieve (Ebbinghaus, 1913; Schacter, 1999). For instance, in a recent study, truth-tellers and liars witnessed a social interaction and then reported it either immediately after or three weeks later (Harvey, Vrij, Leal, Hope, & Mann, 2017). For all of the liars and some of the truth-tellers, the social interaction was important to their task and they intentionally encoded it. For the remaining truth-tellers, the social interaction was unimportant and they incidentally encoded it. Immediately after, truth-tellers who intentionally encoded the interaction reported more details than both liars and truth-tellers who incidentally encoded the interaction. Three weeks later, however, there were no differences in the amount of detail reported by liars and both types of truth-tellers – truth-tellers simply forgot some details, while liars continued to report the same amount of detail. Meanwhile, in a study exploring memory for alibi evidence, people were asked to produce an alibi for three weeks earlier and then spend a week searching for evidence to verify this alibi before retelling their alibi (Strange, Dysart, & Loftus, 2014). People were mostly inconsistent when retelling their alibis, simply because they had not accurately recalled what they did three weeks ago the first time around. Similarly, researchers suggest that a longer time delay might put truthful suspects at risk of forgetting their past activities and making more statement-evidence inconsistencies (Granhag & Hartwig, 2008; Hartwig et al., 2006; Vrij, Granhag, & Porter, 2010) — but this hypothesis has not been tested, until now.

If truth-tellers do make more statement-evidence inconsistencies after a longer time delay, the key issue is that they could be perceived to be lying. People generally perceive

verbal inconsistencies as a sign of deception (Brewer, Potter, Fisher, Bond, & Luszcz, 1999). For example, both police and laypeople are more likely to judge suspects to be guilty when they change their alibis (Culhane & Hosch, 2012). Relatedly, professional lie-catchers such as police, prosecutors, and judges expect truthful statements to be more internally consistent than deceptive statements (Stromwall & Granhag, 2003). It is therefore likely that if truthful suspects are inconsistent with police evidence after a longer time delay, they will appear deceptive, which in turn reduces the diagnostic utility of statement-evidence inconsistencies as a cue to deception.

In two experiments, we explored whether truthful and deceptive suspects contradict evidence to the same extent after an extended time delay. In Experiment 1, subjects were asked to visit a university bookshop and search for a hole-punch (truth-tellers) or steal a wallet from a bag (liars) in a procedure similar to that of Hartwig et al. (2006). The activity was set up to generate evidence, such as eyewitness accounts, that implicated both truth-tellers and liars in the theft of the wallet. Subjects were questioned about their activity in the bookshop either shortly after or two months after without being informed of the evidence implicating them in the theft. Note that in this study, as in past SUE research, all truth-tellers were innocent and all liars were guilty (e.g., Hartwig et al., 2005, 2006). We did not manipulate the factors of guilt and deception separately given that deliberately deceptive innocent suspects are likely to be uncommon in the real world and truth-telling guilty suspects would essentially be confessing to the crime, a scenario in which lie detection would be unnecessary. In Experiment 2, an independent group of laypeople read these subjects' responses and judged to what extent the subjects were lying.

Experiment 1

Method

Subjects and design. A power analysis using G*Power (Faul, Erdfelder, Lang, & Buchner, 2007), assuming a large effect size of $f = 0.4$ (based on Hartwig et al., 2014), and $\alpha = 0.05$ for 4 groups, suggested a sample size of 84 subjects (21 subjects per group) would be sufficient for a power of 0.95. In total, 136 people from the University of Warwick community participated in Phase 1. Six subjects from the truthful condition were excluded at Phase 1 for bringing the wallet to the researcher (n = 3) or for failing to touch the bag while searching for the hole-punch (n = 3, confirmed by a research assistant observing the subject). A further 11 subjects (4 long-delay truth-tellers, 4 long-delay liars, 2 short-delay truth-tellers, and 1 short-delay liar) failed to complete Phase 2, and we excluded one subject from the deceptive condition after Phase 2 for not lying about stealing the wallet. The final sample consisted of 118 subjects (112 provided their age, $M = 20.6$ years, $SD = 3.4$, range = 18–44; 74 women, 42 men, and 2 subjects who identified as other). Subjects were randomly assigned to one of four conditions in a 2 (Suspect: truth-teller vs. liar) x 2 (Time delay: short vs. long) between-subjects design. Cell size ranged from 28–31 subjects. The key dependent variable was subjects' statement-evidence inconsistencies.

Upon completing the experiment, subjects were entered in a lottery to win one of ten £10 vouchers. The research was approved by the University of Warwick's Psychology Research Ethics Committee.

Procedure

Phase 1. The study was advertised online as a real-world search and retrieval study. Subjects participated individually. They met a researcher at the Warwick Arts Centre who instructed them to retrieve an object from the University bookshop, also located within the Arts Centre. Subjects had to enter the bookshop and walk past two potted plants at the entrance. Truth-tellers had to retrieve a hole-punch from inside an orange box in the second row of bookshelves (the Law section), where a computer and telephone are prominently

located. If they could not find the hole-punch after searching for a minute, subjects were to return to the researcher. In fact, there was never a hole-punch in the orange box, so no one found the hole-punch. To search for it, however, truth-tellers had to move a black bag, with a brown wallet sticking out, from on top of the orange box.

Liars were informed that the study would involve committing a fake crime: stealing a wallet from the bookshop. For ethical reasons, subjects were informed that this was not a real crime since the wallet belonged to the researcher and the bookshop's manager and employees were fully aware that the study was taking place. If subjects did not object to stealing the wallet (none did), they visited the second row of bookshelves and searched for the same black bag that truth-tellers searched for and moved. Liars took the wallet from the bag and brought it back to the researcher. Following Phase 1, which took between five to ten minutes, all subjects were told that they would be sent an online questionnaire within a few months. To simulate a real-life criminal investigation in which innocent suspects are unaware that they will be later questioned about their actions, we did not inform subjects what the online questionnaire was about. All subjects were thanked and sent home.

During Phase 1, a research assistant covertly observed subjects from within the bookshop. The research assistant, positioned a few meters away from the orange box and bag, verified whether each subject visited the Law section and handled the bag, either while searching for the hole-punch or while removing the wallet. Thus, Phase 1 generated three pieces of evidence implicating each subject in the theft: (1) eyewitness testimony that the subject entered the bookshop; (2) eyewitness testimony that the subject visited the Law section where the wallet was located; and (3) the subject's fingerprints found on the bag containing the wallet.

Phase 2. Phase 2 took place online to minimize attrition in the long-delay conditions. Short-delay subjects received an online questionnaire on the day they completed Phase 1 and

long-delay subjects received the questionnaire two months after they completed Phase 1. Long-delay subjects who failed to complete the questionnaire in a timely manner were sent a reminder one week later. Short-delay subjects completed Phase 2 within 0–3 days of Phase 1 ($M = 0$, $SD = 0.8$ days) while long-delay subjects completed Phase 2 within 55–80 days of Phase 1 ($M = 63$, $SD = 3.6$ days).

The questionnaire started by asking subjects what they thought was the study's purpose (no one guessed correctly). All subjects were then informed that they were suspected of stealing a wallet from the bookshop. They were told that the investigators had some information indicating their guilt, but as they were not certain, they wanted to question the subject. Truth-tellers were instructed to tell the truth about their Phase 1 activity while liars were instructed to lie about taking the wallet. All subjects were told that to stand a chance to win a £10 voucher, they needed to convince the investigators of their innocence. In fact, all subjects were entered into a lottery to win £10 vouchers regardless of their performance on the task. As a comprehension check, subjects were asked what they were expected to do next and given four options (e.g., "I should lie about what I did in Phase 1 of the study"). Subjects could only proceed with the study when they answered correctly according to their condition. If subjects answered incorrectly, they were prompted to reread the instructions and answer the question again.

Next, subjects were asked SUE-style questions relating to evidence generated in Phase 1, see Table 1. These questions were modelled on interview questions used in past SUE research (Hartwig et al., 2014; Luke, Hartwig, Shamash, & Granhag, 2016). While SUE also involves asking suspects for a free recall account, we only asked specific questions because liars contradict evidence more when answering evidence-specific questions than when freely recalling an account (Hartwig et al., 2011). Each question was presented on a separate page so that subjects could not change their answers to earlier questions. Unlike

face-to-face interviews in which an interviewer can flexibly change their line of questioning based on a suspect's responses, the online questionnaire required subjects to answer a fixed set of questions regardless of how they answered initial questions. For instance, even if a subject claimed they did not visit the bookshop in response to the first question, they were still presented with the remaining questions about their activity within the bookshop. In response, subjects continued to deny visiting the bookshop. All subjects were then asked whether they took anything from the black bag (everyone said no). Subjects were never informed of the evidence implicating them in the theft. Finally, subjects rated the difficulty of the task (1 = *not difficult at all*; 7 = *extremely difficult)*, provided demographic information, and were debriefed.

Results and Discussion

Coding subjects' responses. To measure subjects' statement-evidence inconsistency, two highly trained independent raters, blind to condition and the study's hypotheses, separately coded all subjects' responses. The raters followed Sorochinski et al.'s (2014) coding scheme. For each piece of evidence, the raters decided whether the subject's statements, as a whole (across all 12 questions), were completely consistent with the evidence (score of 1), partially or possibly consistent with the evidence (score of 2), or completely inconsistent with the evidence (score of 3). Although the 12 questions were formulated in relation to specific pieces of evidence (see Table 1), subjects sometimes described their complete activity within the bookshop in response to a single question. Thus, raters considered subjects' responses to all 12 questions when assigning them statement-evidence inconsistency scores for each of the three pieces of evidence.

For an example of how raters coded a subject's statements, consider a subject who responded to the questions by stating that they entered the bookshop, cannot remember whether they visited the Law section (or the second row of bookshelves as some subjects

recalled it), and definitely did not handle a black bag. This subject would receive statement-evidence inconsistency scores of 1, 2, and 3 for evidence pertaining to their bookshop entry (eyewitness testimony), their Law section visit (eyewitness testimony), and their handling of the black bag (fingerprint evidence) respectively. Put simply, each subject received three statement-evidence inconsistency scores – one per piece of evidence. Finally, the raters gave each subject a summed, total statement-evidence inconsistency score that could range from 3 (completely consistent with all three pieces of evidence) to 9 (completely inconsistent with all three pieces of evidence). Thus, in our example above, the subject would receive a total statement-evidence inconsistency score of 6. Note that only one subject made internally inconsistent statements by claiming that they did not visit the Law section of the bookshop but that they did handle the black bag (which was located in the Law section). For this subject, the raters assigned statement-evidence inconsistency scores of 3 (completely inconsistent) for the evidence pertaining to the Law section visit and 1 (completely consistent) for the evidence that they handled the black bag. Overall, the two raters had very high agreement, $\kappa = .85$ [95% CI: 0.80, 0.90], $p < .001$ (Landis & Koch, 1977) and disagreements were resolved via discussion.

Main analysis

Statement-evidence inconsistency. Figure 1 shows mean statement-evidence inconsistency ratings for subjects' responses. In line with past research, liars contradicted the evidence more than truth-tellers did – presumably because liars were denying their connection to the theft of the wallet while truth-tellers were forthcoming about their bookshop activity (e.g., Clemens et al., 2011; Hartwig et al., 2014). Additionally, long-delay subjects contradicted the evidence more than short-delay subjects did, fitting with past research in which people forget details and tell less consistent stories over time (Harvey et al., 2017; Strange et al., 2014). A 2x2 between-subjects ANOVA on subjects' statement-evidence

inconsistency scores revealed main effects of suspect condition, $F(1,114) = 113.99$, $p < .001$, $\eta_p^2 = .500$, and time delay, $F(1,114) = 7.55$, $p = .007$, $\eta_p^2 = .062$, but no interactive effect of suspect condition and time delay, $F(1,114) = .30$, $p = .588$, $\eta_p^2 = .003$. Specifically, liars made more statement-evidence inconsistencies than did truth-tellers ($M_{dec} = 7.18$, $SD_{dec} = 1.55$ vs $M_{tru} = 4.50$, $SD_{tru} = 1.23$, $d = 1.92$, [95% CI: 1.48, 2.36]). There was a large difference between statement-evidence inconsistencies made by liars and truth-tellers shortly after the crime ($M_{diff} = 2.83$ [95% CI: 2.07, 3.58], $d = 2.00$ [95% CI: 1.35, 2.65]), and crucially, two months after the crime ($M_{diff} = 2.55$ [95% CI: 1.88, 3.23], $d = 1.94$ [95% CI: 1.32, 2.56]). Moreover, long-delay subjects contradicted evidence slightly more than short-delay subjects ($M_{lon} = 6.20$, $SD_{lon} = 1.83$ vs $M_{sho} = 5.51$, $SD_{sho} = 2.00$, $d = 0.36$ [95% CI: -0.01, 0.73]). Finally, a follow-up independent samples *t*-test highlighted that short-delay liars made more statement-evidence inconsistencies than did long-delay truth-tellers, $t(57) = 5.21$, $p < .001$, $M_{diff} = 2.00$ [95% CI: 1.23, 2.76], $d = 1.36$ [95% CI: 0.78, 1.92]. Together, these results suggest that truth-tellers were uniformly at less risk of making statement-evidence inconsistencies than were liars.

Inconsistency with individual pieces of evidence. Given that even truth-tellers were contradicting some evidence after both time delays, it would be useful for the police to know which types of evidence truth-tellers and liars might be more likely to contradict. To this end, we examined how many subjects in each condition were completely consistent, partially or possibly consistent, and completely inconsistent with each piece of evidence. There are two key points to note from this analysis, shown in Table 2. First, many truth-tellers (39% short delay; 63% long delay) failed to report handling the bag and contradicted the fingerprint evidence. Unlike entering the bookshop and visiting the Law section, handling the black bag was irrelevant to the truth-tellers' task of finding the hole-punch. Therefore, truth-tellers might have failed to initially notice the bag or later recall it, similar to past research in which

people are less likely to notice or recall things that are irrelevant to their activity (Harvey et al., 2017; Rees et al., 1999; Simons & Chabris, 1999). Indeed, in Harvey et al.'s study, truth-tellers who incidentally encoded the social interaction later reported fewer details than truth-tellers who were instructed to attend to the social interaction as part of their task. Likewise, in this study, task irrelevance at the time of encoding could explain why truth-tellers contradicted the fingerprint evidence at surprisingly high rates, particularly after a long delay. Of course, the police might not be able to assess whether information was relevant to a suspect at the time of encoding without first knowing whether the suspect is lying or telling the truth. Nonetheless, it is important for the police to bear in mind that truthful innocent suspects could also make statement-evidence inconsistencies simply because information which is relevant to the police's evidence was irrelevant to the activity of a truthful innocent suspect.

The second point to note is that liars' contradictions of the evidence became more pronounced the more incriminating the evidence was. While some liars admitted to entering the bookshop, hardly any liars reported handling the black bag that contained the wallet. This result provides further support for liars' counter-interrogation strategy of denial in which they distance themselves from the crime (Granhag & Hartwig, 2008; Hartwig et al., 2014). In this study, the fingerprint evidence was both the most incriminating (leading most liars to contradict it) and the least relevant to truth-tellers' activity (leading many truth-tellers to contradict it). Thus, when using statement-evidence inconsistencies to detect deception, police interviewers may need to consider how incriminating the evidence is as well as its potential irrelevance to truthful suspects.

Difficulty scores. Across conditions subjects found answering interview questions moderately easy ($M = 2.71$, $SD = 1.74$). A 2x2 between-subjects ANOVA showed that overall, suspect condition and time delay had no main or interactive effects on subjects'

perceptions of difficulty, Fs(1, 114) < .01, ps > .95. Thus, surprisingly, subjects did not find it harder to answer interview questions about an activity from two months ago compared to an activity from the last few days.

Overall, Experiment 1 showed that when liars are unaware of the evidence, they contradicted the evidence more than truth-tellers did, even after a two month delay – though a surprisingly high number of truth-tellers contradicted the fingerprint evidence, particularly after a long delay. In Experiment 2, we examined whether laypeople were sensitive to these differences between liars and truth-tellers by presenting mock suspect responses from Experiment 1 to an independent group of subjects and asking them to rate how deceptive they believe the mock suspects are.

Experiment 2

Method

Subjects and design. We presented each mock suspect's response from Experiment 1 to at least two laypeople. A sample of laypeople were recruited via Amazon Mechanical Turk. Only people who judged their fluency in English to be excellent or native could participate. In total, 250 people took part but 13 people were excluded from completing the study after failing the comprehension check. Thus, the final sample consisted of 237 laypeople (M = 38.5 years, SD = 11.9, range = 18–76; 125 men and 112 women).

Laypeople were shown a single mock suspect's response from one of four cells produced by the 2 (Suspect condition: truth-teller vs. liar) x 2 (Time: short delay vs. long delay) between-subjects design. Cell sizes ranged from 56–62 subjects. The dependent variable was laypeople's perceptions of mock suspects' deceptiveness, measured using a forced-choice response and Likert scale. The research was approved by the University of Warwick's Psychology Research Ethics Committee.

Procedure. The procedure consisted of two stages. First, laypeople read some background case information. Then they read the interview questions and one mock suspect's answers from Experiment 1 before judging the mock suspect's deceptiveness. More specifically, laypeople were initially informed that they would be presented with a fictional case of theft and that their task was to judge whether a student being questioned about the theft was telling the truth. We referred to mock suspects as students to prevent laypeople from judging all mock suspect responses as deceptive. Laypeople were shown a photo of the wallet and bag in the bookshop and informed that the wallet had been stolen from the bookshop. Next, laypeople were informed of the evidence from Experiment 1 implicating the student in the theft. We emphasized that the evidence did not prove the student's guilt – only that the student visited the bookshop and handled the bag, possibly accidentally. At this point, laypeople answered a multiple-choice comprehension question about the information they had just read and only laypeople who answered correctly could proceed with the study.

In the second stage, laypeople were informed that the bookshop owner questioned the student about the theft but that the student did not know the evidence. The interview questions and respective mock suspects' answers were shown to laypeople along with a reminder of the evidence. Laypeople were additionally told that there was a 50% chance that they were seeing a response from a truthful student and a 50% chance that they were seeing a response from a deceptive student. We included this instruction about the base rate of truthful and deceptive responses because laypeople might assume that anyone questioned about a crime is guilty (Hartwig et al., 2005; Kassin, Goldstein, & Savitsky, 2003). Laypeople were then asked two questions. First they were asked "Do you think the student is lying?" (yes or no). Next they were asked, "To what extent is the student lying?" (1 = *completely truthful*; 9 = *completely lying*). Finally, all laypeople were asked for demographic information before being thanked and debriefed.

Results and Discussion

Deception judgements. Starting with laypeople's responses to the forced choice question: Laypeople were more likely to indicate that liars were lying than truth-tellers after both time delays. We conducted a between-subjects logistic regression in which the suspect condition and time delay factors were dummy coded, including exploratory analyses on the interaction effect of suspect condition and time delay on people's judgements about whether mock suspects were lying, Wald's $\chi^2(1) = 3.02$, $p = .082$, OR = 3.67 [95% CI: 0.85, 15.92]. Specifically, laypeople perceived more liars to be lying than truth-tellers after a short delay (91.5% vs. 46.4%), Wald's $\chi^2(1) = 21.92$, $p < .001$, OR = 12.46 [95% CI: 4.33, 35.83] and to some extent after a long delay too (90.3% vs. 73.3%), Wald's $\chi^2(1) = 5.54$, $p = .019$, OR = 3.39 [95% CI: 1.23, 9.39]. Laypeople, however, perceived more truth-tellers to be lying after a long delay (73.3%) compared to after a short delay (46.4%), Wald's $\chi^2(1) = 8.49$, $p = .004$, OR = 3.17 [95% CI: 1.46, 6.90]. Laypeople were equally likely to perceive liars to be lying after both a short delay (91.5%) and a long delay (90.3%), Wald's $\chi^2(1) = 0.05$, $p = .818$, OR = 0.86 [95% CI: 0.25, 3.00]. Though laypeople perceived more liars to be lying than truth-tellers after both time delays, they also perceived a surprisingly high number of truth-tellers to be lying, particularly after a long delay. As Table 2 shows, even truth-tellers were not completely consistent with all three pieces of evidence and this could explain why laypeople incorrectly perceived so many truth-tellers to be lying. These results suggest that SUE might assist in detecting deception even with a two month delay between the crime and interview, but there is also a risk that truth-tellers might appear less credible after two months.

Turning now to laypeople's responses on the Likert scale: Figure 2 displays laypeople's mean ratings of the extent to which they thought mock suspects were lying (1 = *completely truthful*; 9 = *completely lying*). Laypeople rated liars as less credible than truth-tellers both immediately after the bookshop visit and two months later, though they also rated

truth-tellers as somewhat less credible after two months. A 2x2 between-subjects ANOVA revealed main effects of mock suspect condition, $F(1, 233) = 50.38$, $p < .001$, $\eta_p^2 = .178$ and time delay, $F(1, 233) = 16.12$, $p < .001$, $\eta_p^2 = .065$, as well as an interaction effect of mock suspect condition and time delay on laypeople's deceptiveness ratings, $F(1, 233) = 4.31$, $p = .039$, $\eta_p^2 = .018$. Laypeople rated liars as more deceptive than truth-tellers, $M_{diff} = 1.94$ [95% CI: 1.38, 2.51], $d = 0.87$ [95% CI: 0.60, 1.14], and mock suspects questioned after a long delay as more deceptive than mock suspects questioned after a short delay, $M_{diff} = 1.09$ [95% CI: 0.48, 1.70], $d = 0.46$ [95% CI: 0.20, 0.72]. Follow-up analyses for the interaction revealed that laypeople rated liars as more deceptive than truth-tellers after both a short time delay, $F(1, 233) = 40.87$, $p < .001$, $M_{diff} = 2.54$ [95% CI: 1.76, 3.32], $d = 1.13$ [95% CI: 0.73, 1.53], and a long time delay, $F(1, 233) = 13.00$, $p < .001$, $M_{diff} = 1.39$ [95% CI: 0.63, 2.15], $d = 0.69$ [95% CI: 0.32, 1.06]. Meanwhile, laypeople rated truth-tellers as more deceptive after a long delay than after a short-delay, $F(1, 233) = 18.16$, $p < .001$, $M_{diff} = 1.69$ [95% CI: 0.91, 2.47], $d = 0.66$ [95% CI: 0.28, 1.04]. Finally, time delay did not impact laypeople's deceptiveness ratings for liars, $F(1, 233) = 1.92$, $p = .167$.

As in Experiment 1, we conducted an independent samples *t*-test[1] comparing laypeople's deceptiveness ratings of short-delay liars and long-delay truth-tellers, $t(105.43) = 2.35$, $p = .020$. Laypeople rated short-delay liars as only slightly more deceptive than long-delay truth-tellers, $M_{diff} = 0.85$ [95% CI: 0.14, 1.57], $d = 0.43$ [95% CI: 0.07, 0.79]. Thus, truth-tellers questioned after two months may be at risk of appearing almost as deceptive as liars questioned immediately after the crime.

Relationship between deception ratings and statement-evidence inconsistencies. Recall that, in Experiment 1, mock suspects received statement-evidence inconsistency scores

[1] Thank you to the anonymous reviewer for suggesting comparisons between these two groups.

between 3 (completely consistent with all evidence) and 9 (completely inconsistent with all evidence). We found that these statement-evidence inconsistency scores positively correlated with laypeople's deception ratings of mock suspects, meaning the more mock suspects contradicted evidence, the more deceptive they appeared to laypeople, $r(235) = .531$ [95% CI: 0.43, 0.62], $p < .001$. Thus, as in past studies, laypeople likely relied upon mock suspects' statement-evidence inconsistencies when making deception judgements (Hartwig et al., 2005, 2006). This correlation, however, does not explain why laypeople rated long-delay truth-tellers as more deceptive than short-delay truth-tellers on the Likert scale despite both groups having similar total statement-evidence inconsistency scores. As Table 2 shows, one key difference between the groups is that long-delay truth-tellers contradicted the fingerprint evidence more than short-delay truth-tellers did. Indeed, laypeople's deception ratings and mock suspects' inconsistency with the fingerprint evidence are positively correlated, $r(235) = .552$ [95% CI: 0.46, 0.63], $p < .001$. Therefore, laypeople might have rated long-delay truth-tellers as more deceptive than their short-delay counterparts because they gave more weight to mock suspects' contradictions of the fingerprint evidence, the most incriminating evidence, than to contradictions of the less incriminating eyewitness evidence.

General Discussion

To our knowledge, this is the first study looking at truthful and deceptive suspects' responses to interview questions after an extended delay of two months. Extending past research on the SUE technique, we found that liars contradicted evidence more than truth-tellers did both shortly after and two months after the crime (Clemens et al., 2011; Hartwig et al., 2005, 2014). The pattern of mock suspects' responses in this study is consistent with past research on liars and truth-tellers' counter-interrogation strategies (Granhag & Hartwig, 2008, 2015). Liars adopted a denial strategy to distance themselves from the stolen wallet and appear credible, while truth-tellers were forthcoming and were more likely to disclose what

they did in the bookshop. Moreover, we found that laypeople, likely relying on statement-evidence inconsistencies, rated liars as more deceptive than truth-tellers after both time delays. Notably, laypeople rated truth-tellers questioned two months after the crime as more deceptive than truth-tellers' questioned shortly after the crime – possibly because many truth-tellers who were questioned two months after the crime contradicted the fingerprint evidence. These findings are consistent with past research in which people infer that someone is being deceptive on the basis of their verbal inconsistencies (Brewer et al., 1999; Culhane & Hosch, 2012; Stromwall & Granhag, 2003).

The finding that truth-tellers are more consistent with evidence than are liars even after a two month delay might appear to conflict with past research in which truthful, innocent mock suspects forget details of their past activities from only three weeks ago (Harvey et al., 2017; Strange et al., 2014). However, truth-tellers in our study likely forgot some details, such as handling the black bag, after the extended delay of two months too, but crucially, they still recalled enough of their past activity, such as visiting the bookshop and its Law section, to respond more consistently with the evidence than did liars. Thus, time delay and memory decay might constrain the effectiveness of some lie detection techniques which rely upon the richness of detail reported by suspects (Harvey et al., 2017), but perhaps not other lie detection techniques, such as SUE which relies upon how much suspects contradict police evidence. The pattern of mock suspects' verbal responses in this study suggest that overall, liars' tendency to distance themselves from the crime outweighs any memory decay truth-tellers might experience in the two months following a crime. In other words, even if truthful suspects forget some details of their activity and contradict a single piece of evidence, they are still more likely to be consistent with the evidence overall than deceptive suspects who deliberately deny their connections to the crime to appear credible.

On a practical level our findings provide further support for the SUE technique by showing that overall, statement-evidence inconsistencies could serve as diagnostic cues to deceit even after an extended time delay and that the police might be able to employ the SUE technique to detect deception effectively even when questioning suspects two months after the crime. Caution is recommended, however, when interpreting a suspect's contradiction of an individual piece of evidence as it might be less indicative of deception and instead reflect a truthful suspect's failure to encode or later recall task-irrelevant information from the time of the crime. This is crucial given that even all of the truth-tellers questioned shortly after the crime did not respond completely consistently with the evidence, and of course, the misclassification of even one truthful suspect as deceptive could have devastating consequences for the accused individual.

In the current study we created an everyday situation—a visit to a bookshop—that innocent, truthful suspects might have to recall during a police interview. Crucially, we did not inform truthful mock suspects that they would be questioned about their activity two months later to prevent them from attending to the activity more closely than an everyday activity or rehearsing their memory of the activity in preparation for the interview. Nonetheless, anecdotally, most of our subjects reported that they had never visited the University bookshop as part of a research study so their experience in this study might have been memorable, making it easier for truth-tellers to respond consistently with evidence even two months later. Future research could test mock suspects' memories for both mundane and novel tasks and locations, to better mirror the activities that truthful suspects might need to recall at interview.

Another matter for future research is testing the effectiveness of the SUE technique with even longer time delays as truthful suspects might be increasingly prone to forgetting the past and contradicting evidence over longer timeframes. The obvious risk is that people

within the criminal justice system, including police interviewers and jurors, might then perceive truthful suspects to be even more deceptive. Relatedly, while laypeople in this study rated short-delay liars as only slightly more deceptive than long-delay truth-tellers, the laypeople were not informed of the time delay between the mock suspect's visit to the bookshop and being questioned about the visit. It is likely that in practice, however, police interviewers and jurors will be aware of the time delay between an alleged incident and a suspect being questioned about the incident. Therefore, it would be useful to explore to what extent police interviewers and jurors adjust their deception judgements of suspects to account for the passage of time and its effect on the number of statement-evidence inconsistencies even an innocent truthful suspect might make.

Finally, we scored mock suspects' statement-evidence inconsistencies using a coding scheme that is typically employed in SUE research (e.g. Sorochinski et al., 2014) to ensure that the results of our study are, to some extent, comparable with the growing number of past SUE studies in which shorter time delays are employed. Nevertheless, more sophisticated coding schemes could be used in future SUE research. For instance, given that mock suspects in this study contradicted different pieces of evidence at different rates depending on how irrelevant or incriminating it was, researchers could differentially weigh statement-evidence inconsistencies based on the importance of the evidence within the case. Moreover, a mock suspect's claim that they do not know or cannot remember something, such as visiting a particular location, is not actually consistent or inconsistent with evidence showing that they visited that location. Thus, an alternative to assigning these ambiguous statements a score of 2 (possibly consistent with the evidence) – which still contributes to the suspects' total statement-evidence inconsistency score – is to simply focus on mock suspects' remaining statements for inconsistencies with the evidence. In other words, researchers could use a binary code and categorize mock suspects' statements as either completely consistent or

completely inconsistent with each piece of evidence. If necessary, suspects' ambiguous statements, which might be more likely after greater time delays, could be treated as a separate measure to statement-evidence inconsistencies.

In conclusion, though researchers suggested that an extended time delay between the crime and interview might make the SUE technique less diagnostic in lie detection, this study has demonstrated that truthful mock suspects might still recall enough information after two months to respond more consistently with the evidence than liars (Granhag & Hartwig, 2008; Hartwig et al., 2006; Vrij et al., 2010). So, even if you can't recall exactly what you were doing two months ago when the police question you – chances are, you will still sound more credible than a liar.

Acknowledgements

We are deeply grateful to Danni Norman, Lottie Devey-Smith, Charlotte Cartwright, Sarah Olin, Dan Gunnell, Lauren Haynes, and Nafeesa Gul for their assistance in collecting and coding the data. We also thank the managers and employees of Warwick University's Bookshop for generously allowing us to conduct our study within the Bookshop during term-time.

References

Association of Chief Police Officers. (2014). *National policing position statement: Pre-interview briefings with legal advisers and information to be supplied to unrepresented detainees.* Retrieved from College of Policing website: http://library.college.police.uk/docs/APPREF/National-Policing-Position-Statement-Pre-Interview-Briefings-2014.pdf

Brewer, N., Potter, R., Fisher, R. P., Bond, N., & Lusczc, M. A. (1999). Beliefs and data on the relationship between consistency and accuracy of eyewitness testimony. *Applied Cognitive Psychology*, *13*, 297–313. doi:10.1002/(SICI)1099-0720(199908)13:43.0.CO;2-S

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

Clemens, F., Granhag, P. A., Strömwall, L. A., Vrij, A., Landström, S., Hjelmsäter, E. R. A., & Hartwig, M. (2010). Skulking around the dinosaur: Eliciting cues to children's deception via strategic disclosure of evidence. *Applied Cognitive Psychology*, *24*, 925–940. doi:10.1002/acp.1597

Colwell, K., Hiscock-Anisman, C., Memon, A., Woods, D., & Michlik, P. M. (2006). Strategies of impression management among deceivers and truth-tellers: how liars attempt to convince. *American Journal of Forensic Psychology*, *24*, 31–38.

Culhane, S. E., & Hosch, H. M. (2012). Changed alibis: Current law enforcement, future law enforcement, and layperson reactions. *Criminal Justice and Behavior*, *39*, 958–977. https://doi.org/10.1177/0093854812438185

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. (2013). Helping to sort the liars from the truth- tellers: The gradual revelation of information during investigative interviews. *Legal and Criminological Psychology*, *20*, 114–128. doi:10.1111/lcrp.12016

Ebbinghaus, H. (1913/1885). *Memory: A Contribution to Experimental Psychology*. Translated by Henry A. Ruger and Clara E. Bussenius. New York: Columbia University.

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

Gilovich, T., Savitsky, K., & Medvec, V. H. (1998). The illusion of transparency: Biased assessments of others' ability to read one's emotional states. *Journal of Personality and Social Psychology*, *75*, 332–346. http://dx.doi.org/10.1037/0022-3514.75.2.332

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

Granhag, P. A., & Hartwig, M. (2015). The Strategic Use of Evidence (SUE) technique: A conceptual overview. In P. A. Granhag, A. Vrij, & B. Verschuere (Eds.), *Deception detection: New challenges and cognitive approaches*. (pp. 231–251). Chichester, UK: John Wiley & Sons.

Granhag, P. A., Rangmar, J., & Strömwall, L. A. (2014). Small cells of suspects: Eliciting cues to deception by strategic interviewing. *Journal of Investigative Psychology and Offender Profiling*, *12*, 127–141. doi:10.1002/jip.1413

Granhag, P. A., Stromwall, L. A., Willen, 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

Hartwig, M., Granhag, P. A., & Luke, T. (2014). Strategic use of evidence during investigative interviews: The state of the science. In D.C. Raskin, C.R. Honts, & J.C. Kircher (Eds.), *Credibility assessment: Scientific research and applications* (pp. 1–36). Oxford, UK: Academic Press.

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

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.A., & Vrij, A. (2005). Detecting deception via strategic disclosure of evidence. *Law and Human Behavior*, *29*, 469–484. doi:10.1007/s10979-005-5521-x

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

Harvey, A. C., Vrij, A., Leal, S., Hope, L., & Mann, S. (2017). Deception and decay: Verbal lie detection as a function of delay and encoding quality. *Journal of Applied Research in Memory and Cognition*, *6*, 306–318. https://doi.org/10.1016/j.jarmac.2017.04.002

Hyman, I. E., Boss, S. M., Wise, B. M., McKenzie, K. E., & Caggiano, J. M. (2010). Did you see the unicycling clown? Inattentional blindness while walking and talking on a cell phone. *Applied Cognitive Psychology*, *24*, 597–607. doi:10.1002/acp.1638

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*, 1–12. 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. http://dx.doi.org/10.1037/0003-066X.60.3.215

Kassin, S. M., Goldstein, C. J., & Savitsky, K. (2003). Behavioral confirmation in the interrogation room: On the dangers of presuming guilt. *Law and Human Behavior*, *27*, 187–203. http://dx.doi.org/10.1023/A:1022599230598

Kassin, S. M., & Norwick, R. J. (2004). Why people waive their Miranda rights: The power of innocence. *Law and Human Behavior*, *28*, 211–221. http://dx.doi.org/10.1023/B:LAHU.0000022323.74584.f5

Landis, J. R., & Koch, G. G. (1977). The measure of observer agreement for categorical data. *Biometrics*, *33*, 159–174. doi:10.2307/2529310

Leins, D., Fisher, R. P., & Vrij, A. (2012). Drawing on liars' lack of cognitive flexibility: Detecting deception through varying report modes. *Applied Cognitive Psychology*, *26*, 601–607. http://dx.doi.org/10.1002/acp.2837

Leins, D., Fisher, R. P., Vrij, A., Leal, S., & Mann, S. (2011). Using sketch-drawing to induce inconsistency in liars. *Legal and Criminological Psychology*, *16*, 253–265. http://dx.doi.org/10.1348/135532510X501775

Lerner, M. J. (1980). *The belief in a just world: A fundamental delusion*. New York, NY: Plenum Press.

Lingwood, J., & Bull, R. (2013). Interviewing young adolescent suspects: When to reveal incriminating information? *The European Journal of Psychology Applied to Legal Context*, *5*, 141–146.

Luke, T. J., Dawson, E., Hartwig, M., & Granhag, P. A. (2014). How awareness of possible evidence induces forthcoming counterinterrogation strategies. *Applied Cognitive Psychology*, *28*, 876–882. http://dx.doi.org/10.1002/acp.3019

Luke, T. J., Hartwig, M., Brimbal, L., Chan, G., Jordan, S., Joseph, E., Osborne, J., & Granhag, P. A. (2013). Interviewing to elicit cues to deception: Improving strategic use of evidence with general-to-specific framing of evidence. *Journal of Police and Criminal Psychology*, *28*, 54–62. doi:10.1007/s11896-012-9113-7.

Luke, T. J., Hartwig, M., Joseph, E., Brimbal, L., Chan, G., Dawson, E., Jordan, S., Donovan, P., & Granhag, P. A. (2016). Training in the Strategic Use of Evidence technique: Improving deception detection accuracy of American law enforcement officers. *Journal of Police and Criminal Psychology*, *31*, 270–278. doi:10.1007/s11896-015-9187-0

Luke, T., Hartwig, M., Shamash, B., & Granhag, P. A. (2016). Countermeasures against the strategic use of evidence technique: Effects on suspects' strategies. *Journal of Investigative Psychology and Offender Profiling*, *13*, 131–147. doi:10.1002/jip.1448

May, L., Granhag, P. A., & Tekin, S. (2017). Interviewing suspects in denial: On how different evidence disclosure modes affect the elicitation of new critical information. *Frontiers in Psychology*, *8,* 1154. https://doi.org/10.3389/fpsyg.2017.01154

McDougall, A. J., & Bull, R. (2015). 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. *Psychology, Crime & Law*, *21*, 1–26. doi:10.1080/1068316X.2014.994631

Rees, G., Russell, C., Frith, C., & Driver, J. (1999). Inattentional blindness versus inattentional amnesia for fixated but ignored words. *Science*, *286*, 2504–2507. doi:10.1126/science.286.5449.2504

Schacter, D. L., (1999). The seven sins of memory: Insights from psychology and cognitive neuroscience. *American Psychologist*, *54*, 182–203, doi:10.1037/0003–066X.54.3.182.

Shawyer, A., Milne, R., & Bull, R. (2009). Investigative interviewing in the UK. In T. Williamson, R. Milne, & S. P. Savage (Eds.), *International developments in investigative interviewing* (pp. 24–38). Cullompton: Willan.

Simons, D. J., & Chabris, C. F. (1999). Gorillas in our midst: sustained inattentional blindness for dynamic events. *Perception*, *28*, 1059–1074. doi:10.1068/p281059

Sorochinski, M., Hartwig, M., Osborne, J., Wilkins, E., Marsh, J., Kazakov, D., & Granhag, P. A. (2014). Interviewing to detect deception: When to disclose the evidence? *Journal of Police and Criminal Psychology*, *29*, 1–8. doi:10.1007/s11896-013-9121-2

Strange, D., Dysart, J. E., & Loftus, E. F. (2014). Why errors in alibis are not necessarily evidence of guilt. *Zeitschrift fur Psychologie*, *222*, 82–89. http://dx.doi.org/10.1027/2151-2604/a000169

Stromwall, 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. http://dx.doi.org/10.1080/10683160308138

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

Sukumar, D., Hodgson, J. S., & Wade, K. A. (2016). How the timing of police evidence disclosure impacts custodial legal advice. *International Journal of Evidence & Proof, 20*, 200–216. doi:10.1177/1365712716643548

Sukumar, D., Wade, K. A, & Hodgson, J. S. (2016). Strategic disclosure of evidence: Perspectives from psychology and law. *Psychology, Public Policy, and Law*, *22*, 306–313. http://dx.doi.org/10.1037/law0000092

Tekin, S., Granhag, P. A., Strömwall, L., Giolla, E. M., Vrij, A., & Hartwig, M. (2015). Interviewing strategically to elicit admissions from guilty suspects. *Law and Human Behavior*, *39*, 244–252. doi:10.1037/lhb0000131

Tekin, S., Granhag, P. A., Stromwall, L. A., & Vrij, A. (2016). How to make perpetrators in denial disclose more information about their crimes. *Psychology, Crime & Law*, *22*, 561–580. http://dx.doi.org/10.1080/1068316X.2016.1168425

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. http://dx.doi.org/10.1007/s10979-008-9143-y

Vrij, A., Granhag, P. A., & Porter, S. B. (2010). Pitfalls and opportunities in nonverbal and verbal lie detection. *Psychological Science in the Public Interest*, *11*, 89–121. http://dx.doi.org/10.1177/1529100610390861

Wachi, T., Kuraishi, H., Watanabe, K., Otsuka, Y., Yokota, K., & Lamb, M. E. (2017). Police officers' ability to detect lies within a deception paradigm. *Psychology, Public Policy, and Law*, *23,* 301–311. http://dx.doi.org/10.1037/law0000124

Table 1.

SUE-style interview questions relating to each piece of evidence.

Eyewitness testimony of subject entering bookshop	Eyewitness testimony of subject visiting second row of shelves	Subject's fingerprints found on the bag containing wallet
Did you visit the bookshop on campus? Did you pass by the two potted plants?	What did you do in the bookshop? Which parts of the bookshop did you visit? Did you see a computer and telephone? Did you visit the Law section of the bookshop?	Did you see a black bag in the bookshop? If yes, where was the black bag? Did you do anything with the black bag? Did you handle the black bag? Did you see anything in the black bag? If yes, what did you see in the black bag?

Table 2.

Percentage of suspects as a function of the consistency of their responses with each piece of evidence.

Evidence	Time delay	Suspect condition	Completely consistent subjects (%)	Partially consistent subjects (%)	Completely inconsistent subjects (%)
Bookshop entry	Short	Truth-teller	100	0	0
		Liar	66	3	31
	Long	Truth-teller	100	0	0
		Liar	58	3	39
Law section visit	Short	Truth-teller	82	14	4
		Liar	21	21	59
	Long	Truth-teller	57	33	10
		Liar	7	23	71
Handling black bag	Short	Truth-teller	54	7	39
		Liar	7	0	93
	Long	Truth-teller	27	10	63
		Liar	0	0	100

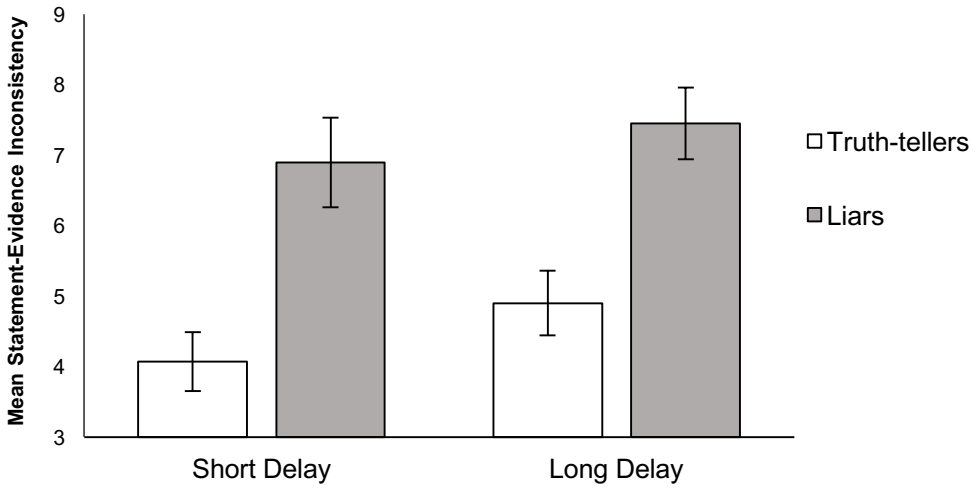

Figure 1. Mean statement-evidence inconsistency ratings of subjects' responses, ranging from 3 (completely consistent with all three pieces of evidence) to 9 (completely inconsistent with all three pieces of evidence). Error bars represent 95% confidence intervals.

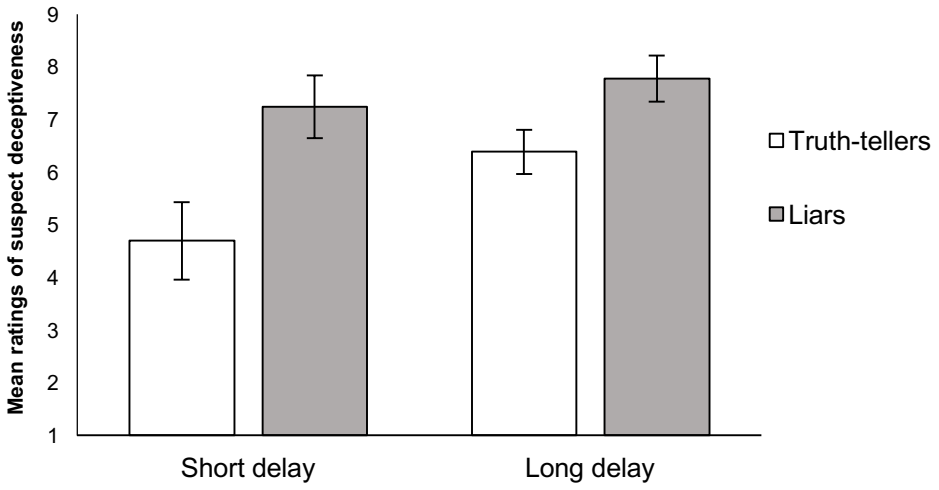

Figure 2. Laypeople's mean ratings of the extent to which the mock suspect was lying (1 = *completely truthful*; 9 = *completely lying*). Error bars represent 95% confidence intervals.

Appendix

Time delay between mock crime and interview in published studies of strategic evidence disclosure

Study	Time delay
Clemens et al. (2010)	Immediately after
Clemens, Granhag, & Stromwall (2011)	Immediately after
Dando & Bull (2011)	1 hour
Dando, Bull, Ormerod, & Sandham (2013)	45 minutes
Granhag, Rangmar, & Stromwall (2014)	Immediately after
Granhag, Stromwall, Willen, & Hartwig (2013)	Immediately after
Hartwig et al. (2011)	10 minutes
Hartwig, Granhag, & Stromwall (2007)	Immediately after
Hartwig, Granhag, Stromwall, & Kronkvist (2006)	Immediately after
Hartwig, Granhag, Stromwall, & Vrij (2005)	1 week
Jordan, Hartwig, Wallace, Dawson, & Xhihani (2012)	Immediately after
Lingwood & Bull (2013)	Immediately after
Luke et al. (2013)	Immediately after
Luke et al. (2016)	Immediately after
Luke, Dawson, Hartwig, & Granhag (2014)	Immediately after
Luke, Hartwig, Shamash, & Granhag (2016)	Immediately after
May, Granhag, & Tekin (2017)	Immediately after
McDougall & Bull (2015)	7–10 days
Sorochinski et al. (2014)	Immediately after
Tekin, Granhag, Stromwall, & Giolla (2015)	Immediately after
Tekin, Granhag, Stromwall, & Vrij (2016)	Immediately after
Wachi et al. (2017)	Immediately after

Note. Some studies gave subjects a few minutes to prepare for the interview.