

Examining the Effects of Evidence Disclosure Timing and Strength on Information Inconsistencies and Provision within Investigative Interviews

Sam Polman¹, Kirk Luther², Hannah de Almeida², Janina Eggers¹, and Steven James Watson^{*1}

¹ University of Twente, NL

² Carleton University, Canada

*Corresponding author information: Dr Steven J. Watson, Department of Conflict, Risk and Safety, University of Twente, Enschede, 7522 NJ, The Netherlands; (e-mail: s.j.watson@utwente.nl).

Abstract:

Late disclosure of evidence within investigative interviews with guilty suspects has been shown to increase statement-evidence and within-statement inconsistencies, which are indicators of untruthful testimony. We experimentally tested whether such inconsistencies were influenced by the timing and strength of the disclosed evidence. We also tested whether evidence disclosure timing or strength had any effect on the provision of novel investigative information or the relationship between interviewer and interviewee, measured via perceived rapport and trust. Participants ($N = 101$) were allocated to one of four groups within a 2(Evidence disclosure timing: Early vs Late) x 2(Evidence strength: Weak vs Strong) between-participants design. Participants received a vignette in which they played a guilty suspect who committed a theft, and were interviewed via videoconference. Suspects were instructed to convince the interviewer that they were innocent. Late disclosure of evidence led to both more statement-evidence inconsistencies and within-statement inconsistencies than early evidence disclosure. In the case of early disclosure, strong evidence produced fewer statement-evidence inconsistencies than weak evidence. Thus, it appears wise to be especially cautious of disclosing strong evidence early. No effects of evidence disclosure timing were found for rapport or novel investigative information. Early disclosure of weak evidence reduced trust in the interviewer, again cautioning against early disclosure. This research supports that late disclosure of evidence can aid lie detection by making use of inconsistencies, and may do so without necessarily impacting on rapport. However, late disclosure did not facilitate the provision of novel investigative information.

Keywords:

Strategic use of evidence; SUE; investigative interview; interrogation; suspect interview; deception detection

Data availability statement:

The data that support the findings of this study are openly available via the Open Science Framework at <https://osf.io/jy7pw/>. We also provide all our data collection materials (vignette, interview script) at this same location.

This study was pre-registered and the (anonymised for peer review) pre-registration is available here: https://aspredicted.org/blind.php?x=/VJK_8X2

Acknowledgements:

We thank Mariëlle Stel for feedback on an earlier draft of this manuscript.

CREDIT

Or:

	Sam Polman	Kirk Luther	Hannah de Almeida	Janina Eggers	Steven J. Watson
Conceptualization					
Methodology					
Formal Analysis					
Investigation					
Writing – Original Draft					
Writing – Review & Editing					
Supervision					
Data Curation					
Project administration					

Examining the Effects of Evidence Disclosure Timing and Strength on Information Inconsistencies and Provision within Investigative Interviews

It is widely acknowledged that people are poor lie detectors, whether those people are trained law enforcement personnel or not (Bond & DePaulo, 2006). Nonetheless, law enforcement personnel must make investigative decisions based upon what they learn from suspects during investigative interviews, as must prosecutors, judges, and juries. Disbelieving true statements made by suspects can lead to unjust convictions (Wells et al., 1998) and falling for deceptive suspect statements can negatively affect legal procedures in the long run (Fawcett, 2012). Thus, it is important to develop methods for discriminating true from false statements (Sakrisvold et al., 2017).

A prominent framework for lie detection within investigative interviews is the Strategic Use of Evidence (SUE) technique (Granhag & Hartwig, 2008, 2015; Hartwig et al., 2005). SUE involves disclosing the evidence held by investigators to the suspect after they provided their account, and it appears to maximize the differences between truth tellers and liars' testimony (Granhag & Hartwig, 2015; Hartwig et al., 2005; Oleszkiewicz & Watson, 2021). Specifically, when evidence is disclosed late compared to early, guilty suspects produce more inconsistencies between their initial statement and the available evidence (statement-evidence inconsistencies). Guilty suspects are also more likely to alter their description of events after evidence is presented (within-statement inconsistencies). Hartwig et al. (2006) trained police trainees to use the SUE technique; their results showed that these officers were more successful at detecting deception than those who were not. Hartwig et al. (2014) meta-analyzed the literature on SUE and concluded that SUE was an effective way of obtaining deceptive cues.

However, a more recent meta-analytic review by Oleszkiewicz and Watson (2021) found that although disclosing the evidence to a suspect later in the interview is likely to be more beneficial in terms of eliciting statement-evidence and within-statement inconsistencies than disclosing evidence early in the interview, the size of this benefit remains unclear. Oleszkiewicz and Watson (2021) stressed that there have been very few recent studies which examine the original form of SUE; that is evidence disclosed early versus late as opposed to variations of gradual disclosure of evidence. The paper by Oleszkiewicz and Watson (2021) also highlighted a clear need for independent verification of the SUE effects. Specifically, of the 15 studies marked as being included in their meta-analyses, 14 included one of the authors of the original Hartwig et al. (2005) experimental demonstration of the SUE effect. The only exception is an unpublished bachelor's thesis (Hingmann, 2019). Therefore, this research

aims to partially replicate, by recreating as far as is possible the methods presented in (Hartwig et al., 2005), the fundamental early versus late disclosure of evidence paradigm in a lab independent of the original authors of SUE. Further, we extend this partial replication by examining whether the effects of early versus late disclosure differ depending upon the strength of the evidence that is disclosed, using a similar design to that employed by Sellers and Kebbell (2009). We also go beyond considering how many cues to deceit are elicited in terms of statement-evidence inconsistencies and within-statement inconsistencies to consider whether evidence disclosure timing or strength has any effect on the number of details elicited from suspects. Finally, we consider whether either the timing or strength of disclosed evidence has any impact on the interpersonal relationship between suspect and interviewer via measures of rapport with and trust in the interviewer.

Suspect counter interview strategies

Both guilty and innocent suspects wish to provide a credible account to convince their interviewer that they are innocent of a crime. However, the theory underpinning SUE argues that innocent and guilty suspects try to achieve this in different ways (Granhag & Hartwig, 2008). Innocent suspects tend to adopt a forthcoming strategy in the hope that this will demonstrate their honesty. Guilty suspects tend to adopt an avoidant strategy (such as not mentioning case relevant information) to prevent revealing incriminating details, and adopt escape strategies (such as silence, refusal to comment, or outright denial) when their account is directly challenged – as long as they believe interviewers do not have evidence that would catch them in a lie when making a denial. Consequently, it is critical for guilty suspects to anticipate what evidence the police might hold against them so that they can make sensible decisions about what incriminating details they can deny or avoid mentioning.

The concepts of information and decision control are key to understanding how suspects make decisions about what to admit, avoid, or deny (Granhag & Hartwig, 2008). Information and decision control are two cognitive self-regulatory strategies that it is argued suspects use to manage the potential threats an investigative interview poses. Suspects gain information control from predicting what might happen in the interview so that they can prepare a counter response. Critical to information control is a prediction of what evidence the investigators may hold against the suspect. Decision control comes from deciding how to act during an interview and it is informed by the process of information control. Guilty suspects are assumed to make decisions about what to avoid, deny, or admit depending on what evidence they believe the investigators are likely to hold (Hartwig et al., 2007). Consequently, guilty suspects are expected to only be forthcoming about details where they believe there is

no harm in making an admission because it does not implicate them in the crime, or where they believe there is no point withholding or denying because the interviewer is likely to hold evidence that would make such evasions pointless or damaging to the credibility of their narrative. Innocent suspects' decision control is expected to be more heavily informed by cognitive biases such as *belief in a just world* – that people get what they deserve and so they will not be considered to be guilty when innocent – and the *illusion of transparency* – that people are able to accurately infer others' states and so the interviewer will be able to see the suspect's innocence (Hartwig et al., 2007). The effect of these cognitive biases is that innocent suspects usually adopt a forthcoming approach to the interview.

The SUE approach is designed to exploit these differences between guilty and innocent suspect strategies to maximize cues to deception in the form of inconsistencies; both within guilty suspect statements, and with the available evidence. This is achieved by manipulating the timing in which evidence is revealed to suspects during an interview.

Evidence disclosure timing and inconsistencies

Broadly, there are three general approaches to disclosing evidence: early – before an account is taken from a suspect, late – after a full account is taken from the suspect, or gradually – in a drip feed fashion depending on when investigators believe it is appropriate to release specific pieces of information (Oleszkiewicz & Watson, 2021). Police interviewers use all three approaches in practice, though with a preference for gradual or late disclosure (Walsh et al., 2016); early disclosure is used by interviewers when they hope a suspect will then confess (Dando & Bull, 2011). Proponents of SUE argue deferred approaches to evidence disclosure (gradual or late) are superior to early disclosure because they prevent suspects changing their accounts to explain already revealed evidence (Granhag & Hartwig, 2008).

The net result of these different approaches to an interview between guilty and innocent suspects (considering the differing interview strategies of guilty and innocent suspects) is that innocent suspects are expected to explain evidence against them even before this evidence is presented to them, while guilty suspects are likely to omit incriminating details or deny them when directly challenged. Therefore, when evidence is presented later in the interview this evidence should more often correspond to the account given by an innocent suspect, while it should contradict either directly or by omission the account given by a guilty suspect. These direct contradictions are known as *statement-evidence inconsistencies*. A second key inconsistency is that the presentation of evidence should provoke a change in a guilty person's story to account for these inconsistencies, while fewer changes in the story of

an innocent suspect's account should be necessary. These changes in an account are named *within-statement inconsistencies*.

Guilty suspects are more likely to make statement-evidence inconsistencies regardless of the timing of evidence disclosure (Hartwig et al., 2014; Oleszkiewicz & Watson, 2021). However, the number of these contradictions greatly increases when evidence is disclosed to guilty suspects late in the interview, but does not increase when innocent suspects have evidence presented early versus late. Thus, when a guilty suspect is not aware of the available evidence against them, they are more likely to make statements that contradict the evidence (Hartwig et al., 2014). Therefore, the late disclosure of evidence appears to be a reliable cue to deception.

Evidence disclosure timing and novel investigative information

Historically, the SUE framework has focused on lie detection (Granhag & Hartwig, 2015). However, the primary objective of an investigative interview is the gathering of information relevant to the crime and novel investigative information (Tekin et al., 2016). Novel investigative information refers to information the interviewer was unaware of prior to interview which can link a suspect to an offense (Tekin et al., 2015). Novel investigative information has gained popularity in research regarding evidence disclosure because it is of greater practical use to practitioners than pure deception detection (Tekin et al., 2016; Walsh & Bull, 2015).

The elicitation of novel investigative information may also be affected by manipulating the timing of the evidence disclosure. However, the direction of any effects remains unclear. According to Oleszkiewicz and Watson (2021), no studies have directly compared the potential effect of late versus early disclosure on the elicitation of novel investigative information. This lack of data reflects that incorporating novel investigative information as a measure emerged at a time when SUE studies began testing gradual disclosure of evidence to encourage suspects to shift their strategy from avoidant towards forthcoming by manipulating a suspect's information control. Briefly, the logic is that the gradual disclosure of information encourages a suspect to re-evaluate their prior assumptions about how much information the interviewer holds. The result of this reappraisal is that suspects are more likely to provide more details about their involvement in a crime in order to attempt to better match any evidence the interviewer may hold in order to avoid contradicting evidence and thus maintain credibility (Luke & Granhag, 2022).

The effect of late (versus early) disclosure of evidence on the elicitation of information is unknown, but we expect late disclosure of evidence is more likely to elicit a more

forthcoming account. Immediately disclosing evidence allows suspects to know exactly what information is held by an interviewer, and so we would expect a guilty suspect that does not intend to confess would likely only reveal information that accounts for revealed evidence. With late disclosure, the evidence held by the interviewer is unknown; thus, suspects may assume the interviewer holds more information than they do and provide more details than would be necessary to explain their activities. Similarly, late disclosure is expected to generate statement-evidence inconsistencies because Tekin et al. (2016) found suggestive evidence that suspects may become more forthcoming when they are confronted with these statement-evidence inconsistencies. Thus, there is a need to directly test how late (versus early) disclosure of evidence might affect how much information is provided by suspects, especially since we know both methods are used in practice (Dando & Bull, 2011; Walsh et al., 2016).

Strength of evidence

Data from investigative interviews revealed that stronger evidence is associated with a greater number of admissions by suspects, especially suspects with limited criminal histories (Cassell & Hayman, 1995; Moston et al., 1992). However, there has been surprisingly limited explorations of the impact of evidence strength on suspect behavior.

In one of the few current explorations of the impact of evidence strength on suspect behavior, Brimbal and Luke (in press) found that statements made by participants about mock crimes were more aligned with the evidence when the evidence was very incriminating. More specifically, they found that more reliable evidence (e.g., evidence from CCTV versus an eyewitness) generated more consistent accounts, while evidence proximity (e.g., placing a suspect in a specific location vs only the general area of a crime) did not affect consistency. Brimbal and Luke argued that more reliable evidence strongly links the suspect to the crime compared to less reliable evidence; consequently, suspects are less inclined to force their stories to account for weaker evidence and so are willing to generate a statement-evidence inconsistency.

Strong evidence disclosed early should lead to few statement-evidence or within-statement inconsistencies, since both early (versus late) and strong (versus weak) evidence provoke fewer inconsistencies. However, early disclosure of weak evidence may provoke more statement-evidence inconsistencies as suspects may feel emboldened to contradict weak evidence. Within-statement inconsistencies should remain low since there is no new evidence presented to provoke contradictions.

The effects with late disclosure may be very different. Late disclosure should generate statement evidence inconsistencies regardless of evidence strength. However, within-statement inconsistencies should be more likely to be provoked by stronger evidence, because suspects may feel less inclined to explain away weak evidence.

We also consider the effect of evidence strength on the elicitation of novel investigative information. Earlier, we argued that suspects provide novel investigative information to ward off inconsistencies between evidence and their statements (Granhag et al., 2009). Following similar logic, we would expect suspects would only be motivated to provide additional information in response to strong evidence regardless of disclosure timing.

Evidence disclosure and interviewer-interviewee relationships

One possible downside of late evidence disclosure is that it could negatively affect the relationship between interviewer and interviewee. Clemens et al. (2020) surveyed German police officers regarding the SUE technique, and one concern raised was that SUE may interfere with the quality of interaction between interviewer and interviewee. That is, SUE may reduce rapport. Given the central importance of rapport in eliciting information (Gabbert et al., 2021), any such negative effect could prove detrimental to the overall objectives of any interview. Moreover, there is the possibility suspects feel deceived and so plausibly late disclosure of evidence may also negatively affect how far the interviewee trusts the interviewer. Any violation of trust may also inhibit the willingness of suspects to provide information because trust has been shown to be associated with generating cooperation and eliciting information from suspects (Collins et al., 2002; Griffiths & Rachlew, 2018). Consequently, we test for any effects of evidence disclosure timing on these interpersonal factors, and test if any effects depend on whether the disclosed evidence is strong or weak.

The Current Study

We build on Brimbal and Luke's work (in press) by adding evidence disclosure timing, to determine which of these effects (timing vs strength of evidence disclosure) predominate in the generation of statement-evidence and within-statement inconsistencies, or if they interact with one another. Our design follows that performed by Sellers and Kebbell (2009), who also manipulated both evidence strength and disclosure timing, to show that late disclosure of evidence increased the rate of confessions, unless weak evidence was presented in which case confessions became less likely. However, we modified our design to address some limitations of Sellers and Kebbell's study (2009) that are also highlighted by Brimbal and Luke (in press). First, Sellers and Kebbell (2009) used the same form of evidence to manipulate evidence strength (eyewitness evidence) and so it might be better argued that they

in truth manipulated evidence accuracy rather than evidence strength. We addressed this by manipulating evidence type so that evidence either more closely ties a suspect to a crime (proximity manipulation, e.g., placing a suspect in a general area versus directly at the scene of a crime) or else is more reliable (e.g., eyewitness versus CCTV footage). Second, Sellers and Kebbell (2009) were primarily interested in the elicitation of confessions and did not measure outcome variables such as the number of inconsistencies elicited from suspects, and novel investigative information elicited.

Summary and Hypotheses

All our hypotheses and predictions refer only to the behavior of guilty suspects. We do not examine the behavior of innocent suspects.

Because early disclosure provides guilty suspects an opportunity to integrate available evidence into their account and facilitates the manufacturing of credible lies, it was hypothesized that early disclosure will lead to fewer statement-evidence and within-statement inconsistencies than late disclosure.

Because strong evidence was found to lead to suspects making statements more consistent with the evidence (Brimbal & Luke, in press), it was hypothesized that strong evidence leads to fewer statement-evidence inconsistencies than weak evidence in the case of early disclosure. In the case of late disclosure, strong evidence should lead to more within-statement inconsistencies than weak evidence because suspects are motivated to account for contradictions between only strong but not weak evidence and their earlier statements.

We further tested whether the disclosure of strong evidence provokes suspects to provide more novel investigative information than weak evidence, because guilty suspects are motivated to account for evidence that implicates them.

Finally, we tested whether the evidence strength or disclosure timing affects the perceived rapport or trust with the interviewer from the suspects perspective.

Our hypotheses are pre-registered via AsPredicted here:

https://aspredicted.org/blind.php?x=/VJK_8X2.

Methods

Participants

Participants were recruited through the participant recruitment system of a Dutch University, and through the personal contacts of the researchers, with 101 participants in total. Our participants were 54.5% were female ($N = 55$) and 45.5% were male ($N = 46$), between 18 and 63 years of age ($M = 25.55$, $SD = 7.34$) and 44.6% were students at the host university

($N = 45$), 10.9% were a student somewhere a different university ($N = 11$), and 44.6% were not a student ($N = 45$).

Design

The study consisted of a 2 (Evidence Disclosure Timing: Early vs Late) x 2 (Evidence Disclosure Strength: Weak vs Strong) between-subjects design. Participants were randomly assigned to one of the four groups. There were 26 participants in the Early Disclosure of Weak Evidence group, and 25 in all other experimental groups.

Procedure and materials

All our materials are available at the project Open Science Framework project page here: **Steven still needs to do this**. This project took place entirely online using vignettes and video calls rather than mock crimes and in-person interviews because data collection took place during a covid-19 lockdown, which made in person data collection or crime simulation impossible.

Pre-interview. The project received ethical approval from the host institution (Approval number: 210103). Two authors were directly involved in data collection and participant recruitment (SP and JE). After signing up to take part in the research, an instruction sheet was sent to the participants informing them that they were required to portray suspects guilty of a theft and that their task was to convince the interviewing officer of their innocence. The crime in question concerned the theft of Apple AirPods from a bag at a library. The participants were asked to imagine that they were short on money. They entered the library to search for a new job when a wealthy-looking man next to them spilled his coffee over their notes after which he left to go to the bathroom. The participants were told they then looked into his bag, saw a sealed, brand-new pair of Apple AirPods and took them. They then went to a store to sell them in return for a voucher.

Interview. Recruited participants were invited to a video call with the research team. The relevant authors (SP and JE) played one of two roles during the interview; they either played the role of researcher (in which they took care of procedural formalities such as preparing the participant and guiding them through the post hoc measures and debrief), or the role of interviewer. If any of the participants were familiar with one of the two researchers, the interviewer was portrayed by the researcher unknown to the participant. During the interview, the researcher muted and turned off their camera off so that only the interviewee (participant) and interviewer were active participants in the chat.

The interviews were based on an interrogation paradigm first introduced by Kassin and Fong (1999) as used by Hartwig et al. (2005) in the initial SUE experiment. The

interviews in the Early Disclosure group began with the researcher introducing themselves as an officer and informing participants they were suspected of having stolen Apple AirPods from a table at the nearby library. Participants were then asked whether they had committed the crime. Next, participants were then presented with the evidence against them. The available evidence in question consisted of the following pieces: (i) in the Weak Evidence group, participants were told that they were logged into a computer in the library at the moment of the theft (low proximity). Furthermore, they were told that their fingerprints were found on the victim's table (low proximity). Lastly, they were told that the receptionist thought she saw someone that looked like the suspect looking in the victim's bag (low reliability). (ii) In the Strong Evidence group, participants were told that they were logged into the computer next to the victim's table at the moment of the theft (high proximity). Moreover, they were told that their fingerprints were found on the victim's bag (high proximity) and that there is camera footage of them going through the victim's bag (high reliability). Participants were then requested to provide the interviewer with a full account of what they were doing during the time of the crime. Afterwards, they were asked whether they had been in the library, whether they had seen the Apple AirPods, and whether they had touched them. At the end of the interview, they were asked one last time whether they confess to the crime or not after which they were thanked for their participation. The participants in the Late Disclosure group were also presented with these questions, but the available evidence was only disclosed to the participants after they finished their initial account of their actions at the time of the crime and answered all questions. After the evidence was disclosed, they were asked to comment on the evidence, whether they confess to the crime, after which the interview was ended.

Post-interview. After the interviews were finished, the interviewer muted themselves and turned the camera off. The researcher who welcomed the participants asked the participant to complete the online post hoc Qualtrics questionnaire. The questionnaire included demographic questions regarding their age and occupation and the interviewee version of the rapport scale for investigative interviews developed by Duke et al. (2018). This scale consisted of 21 statements (e.g., "The interviewer respects my knowledge" and "The interviewer was attentive to me") to which the participants could state to what extent they agreed using a 5-point Likert scale ($1 = strongly disagree, 5 = strongly agree$). High scores indicate high rapport¹. The scale had a good internal consistency ($\alpha = 0.94$).

¹ Note that this scale includes several items which overlap with our measure of trust because they assess, in our opinion, the perceived competence and integrity of the interviewer. We also ran all our analyses using the

Participants also completed a trust scale based on Mayer et al. (1995). The scale was composed of the three underlying trust subscales, which were ability (six items; e.g., “The interviewer seemed to be well qualified”), benevolence (five items; e.g., “I got the impression the interviewer would go out of their way to help me”), and integrity (six items; e.g., “Sound principles seemed to guide the interviewer’s behavior”). Trust was indicated on a 5-point Likert scale (anchored at 1 = *strongly disagree*, and 5 = *strongly agree*) with high scores indicated higher trust. The internal consistency of all test items was good (Cronbach’s α .95).

Interview coding. The recorded interviews were coded by two authors (SP and HdA). Statement-evidence inconsistencies were coded if participants made a statement that either contradicted or failed to account for one of the three pieces of evidence the interviewer held against the suspect (e.g., ‘I was not at the library’ is an inconsistency because the evidence shows they were logged onto a library computer during the time of the crime).

Within-statement inconsistencies were coded when participants made a remark that contradicted something they had stated previously in the interview, but only where this related to one of the three pieces of evidence (e.g., ‘I had already left the library during the time of the crime’ vs ‘Oh yeah, I came back later so maybe I was there’).

Novel investigative information was coded whenever the participants disclosed any checkable statements about what happened during the crime which were not related to the pieces of evidence held by the interviewer.

We assessed interrater agreement on the number of statement-evidence and within-statement inconsistencies using Cohen’s Kappa. We found good agreement for the number of statement evidence inconsistencies ($\kappa = .75$), though agreement was lower for within-statement inconsistencies ($\kappa = .66$). We assess reliability of the rating for the number of Novel Investigative Information with a Pearson correlation given that we had two raters, which indicated a strong association between scores of the two raters ($r = .78$). For robustness, and because we base our analyses on the scores generated by the first author (SP), we also estimated reliability using an Intraclass correlation coefficient with a two-way random model assessing absolute agreement which indicated moderate agreement ($ICC = .68$).

rapport measure with these items excluded. It did not change our results from those we found using the full rapport scale. We include this measure of rapport that omits the items that overlap with trust in our open data. Omitted items were: “I think the Interviewer is generally honest with me.”, “The Interviewer did his/her job with skill during the interview.”, “The Interviewer performed expertly during the interview.”, “I think that the Interviewer can generally be trusted to keep his/her word.”, “I feel I can trust the Interviewer to keep his/her word to me.”, “The Interviewer made an effort to do a good job” and “The Interviewer acted like a professional.”

Results

Table 1 presents the means, standard deviations, and correlations between Statement-Evidence Inconsistencies, Within-Statement Inconsistencies, Novel Investigative Information, Rapport, and Trust.

Statement-Evidence Inconsistencies and Within-Statement Inconsistencies were positively correlated, $r(99) = .46, p < .001$, while Statement-Evidence Inconsistencies and Novel Investigative Information were negatively correlated, $r(99) = -.33, p = .001$. In contrast, Within-Statement Inconsistencies and Novel Investigative Information were positively correlated, $r(99) = .28, p = .005$. Rapport and Novel Investigative Information were not significantly correlated, $r(99) = .08, p = .45$.

Table 1

Descriptive Statistics and Correlations

	M	SD	Rapport	Trust	SEI	WSI
Rapport	3.90	0.74				
Trust	3.29	0.89	.88***			
SEI	1.43	0.91	-.06	-.03		
WSI	0.58	0.83	.02	.05	.46***	
NII	2.21	2.06	.08	.11	-.33***	.28**

$N=101$, SEI = Statement-evidence inconsistency, WSI = within statement inconsistency, NII = Novel Investigative information

* $p < .05$, ** $p < .01$, *** $p < .001$

Before performing statistical tests of our hypotheses, we describe the distribution of our inconsistency outcomes to offer a fine-grained insight to our data and the conditions under which different types of inconsistency are elicited. Table 2 shows the number of statement-evidence inconsistencies that were provided by the suspects in each experimental group. In total, only two people went against all three pieces of evidence by either stating they were not present at the library or by saying nothing at all. In the early disclosure of weak evidence category, 50% went against two pieces of evidence, and the others went against either one or zero pieces of evidence.

In the early disclosure of strong evidence category, 72% provided stories coherent with all pieces of evidence. In both late disclosure categories (weak and strong evidence),

almost all suspects made statements that contradicted two pieces of evidence. Almost all participants placed themselves at the scene of the crime by admitting being at the library, but failed to provide a plausible explanation for the other two pieces of evidence (their fingerprints being found on the victim's table/bag and being spotted by the receptionist/CCTV camera looking into the victim's bag) before being presented with them. Broadly, these findings support our prediction that early disclosure is associated with fewer statement-evidence inconsistencies, especially when there is strong evidence.

It is very important to note that only 1.4% of statement-evidence inconsistencies consisted of fabrications that directly contradicted the evidence and 98.6% of statement-evidence inconsistencies were omissions where the suspect did not mention anything to contradict or account for the available evidence.

Table 2

Frequencies of Statement-Evidence Inconsistencies

Evidence Strength	No. Statement-Evidence Inconsistencies	Frequency (%)	
		Early Disclosure	Late Disclosure
Weak	0	7 (27%)	1 (4%)
	1	6 (23%)	0 (0%)
	2	13 (50%)	23 (92%)
	3	0 (0%)	1 (4%)
Strong	0	18 (72%)	1 (4%)
	1	0 (0%)	0 (0%)
	2	6 (24%)	24 (96%)
	3	1 (4%)	0 (0%)

Table 3 shows the number of within-statement inconsistencies provided by the suspects in each category. In both early disclosure categories (weak and strong evidence), almost no one made any within-statement inconsistencies. In the late disclosure of weak evidence category, most people made one within-statement inconsistency, whereas the others made either zero or two. In the case of late disclosure of strong evidence, most of the participants made either zero or two within-statement inconsistencies, indicating that the disclosure of evidence evoked divergent strategies (the participants stopped actively trying to convince the officer of their innocence and became silent or they tried harder to change their

story into a plausible account). These results appear to support our prediction that there are more within-statement inconsistencies with late versus early disclosure, but support for our prediction that strong evidence would provoke more statement evidence inconsistencies than weak evidence in only indicative.

Table 3

Frequencies of Within-Statement Inconsistencies

Evidence Strength	No. Within-Statement Inconsistencies	Frequency (%)	
		Early Disclosure	Late Disclosure
Weak	0	24 (92%)	7 (28%)
	1	1 (4%)	10 (40%)
	2	1 (4%)	8 (32%)
	3	0 (0%)	0 (0%)
Strong	0	24 (96%)	9 (36%)
	1	0 (0%)	4 (16%)
	2	1 (4%)	12 (48%)
	3	0 (0%)	0 (0%)

Hypothesis testing

We used a factorial ANOVA comparing Disclosure Timing (early versus late) and Strength of Evidence (weak versus strong) for each of our dependent variables.

Statement-Evidence Inconsistencies. We observed statistically significant main effects for both Evidence Disclosure Timing, $F(1, 97) = 50.0, p < .001, \eta_p^2 = .34, 95\%CI [-1.79, -0.93]$, and Evidence Disclosure Strength, $F(1, 97) = 5.36, p = .02, \eta^2 = .05, 95\%CI [-0.76, 0.03]$, on the number of statement-evidence inconsistencies generated. These main effects indicated more Statement-Evidence Inconsistencies for those to whom evidence was disclosed late ($M = 1.94, SD = 0.42$) versus those who were presented with the evidence early ($M = 0.92, SD = 0.98$), $d = 1.35$ (95%CI [0.91, 1.78]). We also found more Statement-Evidence Inconsistencies for those presented with weak ($M = 1.59, SD = 0.78$) versus strong evidence ($M = 1.26, SD = 1.01$), $d = 0.37$ (95%CI [-0.03, 0.76]).

We also observed an interaction effect between Disclosure Timing and Evidence Strength, $F(1, 97) = 4.16, p = .04, \eta_p^2 = .04$. We followed up this interaction with tests of simple effects. This showed that in the case of Late Disclosure, there was no significant effect of Weak ($M = 1.96, SD = 0.45$) versus Strong Evidence ($M = 1.92, SD = 0.40$) on the amount

of Statement-Evidence Inconsistencies, $F(1,97) = 0.04, p = .85$, Mean difference[95%CI] = 0.04 [-0.37, 0.45], $d = 0.09$ (95%CI[-0.46, 0.65]). In the case of Early Disclosure, there was a significant effect of Weak ($M = 1.23, SD = 0.86$) versus Strong Evidence ($M = .60, SD = 1.00$) on the number of Statement-Evidence Inconsistencies, and the size of the effect was medium, $F(1,97) = 5.07, p < .01$, Mean difference[95%CI] = .63 [.23, 1.04], $d = 0.68$ (95%CI[0.11, 1.24]). Thus, as predicted, only in the case of early disclosure does strong evidence produce fewer statement-evidence inconsistencies than weak evidence.

Our tests of simple effects also showed that in the case of Weak Evidence, there was a significant effect of Early ($M = 1.23, SD = 0.86$) versus Late ($M = 1.96, SD = 0.45$) Disclosure on the amount of Statement-Evidence Inconsistencies, $F(1,97) = 12.79, p < .01$, Mean difference[95%CI] = 0.73[0.32, 1.13], $d = -1.06$ (95%CI[-1.64, -0.47]). There was also a significant difference between disclosure timing with Strong Evidence, between Early ($M = 0.60, SD = 1.00$) versus Late ($M = 1.92, SD = 0.40$) Disclosure on the amount of Statement-Evidence Inconsistencies, $F(1,97) = 21.78, p < .01$, Mean difference[95%CI] = 1.32[0.91, 1.73], $d = -1.73$ (95%CI[-1.64, -0.47]). Thus, in line with our hypotheses the case of both weak evidence and strong evidence, late disclosure produced significantly more statement-evidence inconsistencies than early disclosure.

In summary, disclosing the evidence late produced more statement-evidence inconsistencies than disclosing the evidence early, both when the evidence is weak and strong. Only when evidence was disclosed early did strong evidence produce fewer statement-evidence inconsistencies than weak evidence.

Within-Statement Inconsistencies. Regarding Within-Statement Inconsistencies, we observed statistically significant main effects for Disclosure Timing, $F(1, 97) = 53.58, p < .01$, $\eta_p^2 = .36$, 95%CI[1.03, 1.91], but not for Evidence Strength, $F(1, 97) = .03, p = .87, \eta^2 = .00$, 95%CI[-0.35, 0.43]. These main effects indicated more Within-Statement Inconsistencies for those to whom evidence was disclosed late ($M = 1.08, SD = 0.85$) versus those who were presented with the evidence early ($M = 0.10, SD = 0.41$), $d = 1.47$ (95%CI[1.56, 2.06]). There was no significant difference in Within-Statement Inconsistencies regarding Weak ($M = 0.57, SD = 0.78$) versus Strong evidence ($M = 0.60, SD = 0.88$), $d = -0.04$ (95%CI[-0.43, 0.35]). We also observed no interaction effect between Disclosure Timing and Evidence Strength, $F(1, 97) = .19, p = .67, \eta_p^2 = .00$. This indicates that suspects who were presented with weak evidence late ($M = 1.04, SD = 0.79$) and strong evidence late ($M = 1.12, SD = 0.93$) made more within-statement inconsistencies when compared to those who were presented with

weak evidence early ($M = 0.12$, $SD = 0.43$) or those who were presented with strong evidence early ($M = 0.08$, $SD = 0.40$).

These findings support our prediction that there would be more within-statement inconsistencies when evidence is presented late, but not that the number of within-statement inconsistencies would increase to a greater extent when evidence is stronger because participants would be more motivated to correct contradictions to strong over weak evidence.

Novel Investigative Information. We observed no statistically significant main effects for either Disclosure Timing, $F(1, 97) = .05$, $p = .82$, $\eta_p^2 = .00$, 95%CI[-0.43, 0.35], or Evidence Strength, $F(1, 97) = .82$, $p = .37$, $\eta^2 = .01$, 95%CI[-0.57, 0.21], on the elicitation of novel investigative information. We followed up these null findings with an exploratory Bayesian ANOVA, using a default r scale fixed effects of .5, revealed moderate evidence in favour of the null hypotheses for Disclosure Timing ($BF_{01} = 3.30$) and Evidence Strength ($BF_{01} = 4.65$). Therefore, these main effects indicated no difference in the disclosure of novel investigative information between those to whom evidence was presented late vs. early ($M = 2.16$, $SD = 2.05$ vs. $M = 2.25$, $SD = 2.08$, $d = 0.05$, 95%CI[-0.43, 0.35]) nor when the evidence was weak vs strong ($M = 2.39$, $SD = 2.17$ vs. $M = 2.02$, $SD = 1.93$, $d = 0.18$, 95%CI[-0.21, 0.57]).

We also observed no interaction effect between Disclosure Timing and Evidence Strength, $F(1, 97) = .21$, $p = .65$, $\eta_p^2 = .00$, with an exploratory Bayesian ANOVA showing very strong evidence in favour of the null hypothesis $BF_{01} = 50.93$. This indicates there was no significant difference in the amount of novel investigative information being provided by suspects who were presented with weak evidence early ($M = 2.35$, $SD = 2.45$), strong evidence early ($M = 2.16$, $SD = 1.65$), weak evidence late ($M = 2.44$, $SD = 1.89$) and strong evidence late ($M = 1.88$, $SD = 2.20$). Therefore, we found no support for our predictions that stronger evidence would provoke the participants to provide novel investigative information.

Rapport. A factorial ANOVA was conducted to test for effects of Disclosure Timing (early vs late) and Strength of Evidence (weak vs strong) on Rapport. We observed no statistically significant main effects for either Disclosure Timing, $F(1, 97) = 0.27$, $p = .61$, $\eta_p^2 = .00$, 95%CI[-0.28, 0.50], or Evidence Strength, $F(1, 97) = 0.01$, $p = .93$, $\eta^2 = .00$, 95%CI[-0.40, 0.38]. An exploratory Bayesian ANOVA revealed moderate evidence in favor of the null hypotheses for Disclosure Timing ($BF_{01} = 4.20$) and Strength of Evidence ($BF_{01} = 4.76$). Therefore, we found no difference in the perceived rapport between those to whom evidence was presented late vs early ($M = 3.94$, $SD = 0.70$ vs. $M = 3.86$, $SD = 0.78$, $d = 0.11$, 95%CI[-0.28, 0.50]), nor weak vs strong evidence ($M = 3.90$, $SD = 0.75$ vs. $M = 3.89$, $SD = 0.73$, $d =$

0.01, 95%CI[-0.38, 0.40]). We also observed no significant interaction effect between Disclosure Timing and Evidence Strength, $F(1, 97) = 1.67, p = .20, \eta_p^2 = .02$, with an exploratory Bayesian ANOVA showing very strong evidence in favour of the null hypothesis $BF_{01} = 37.80$. This indicates there is no significant difference in the perceived rapport for suspects who were presented with weak evidence early ($M = 3.77, SD = 0.76$), strong evidence early ($M = 3.95, SD = 0.81$), weak evidence late ($M = 4.04, SD = 0.74$) and strong evidence late ($M = 3.83, SD = 0.65$). In summary, we did not find evidence that timing or strength of evidence disclosure or strength had meaningful impact on the level of self-reported established rapport.

Trust. We observed no statistically significant main effects for either Disclosure Timing, $F(1, 97) = 0.32, p = .58, \eta_p^2 < .01, 95\%CI[-0.28, 0.50]$, or Evidence Strength, $F(1, 97) = .03, p = .86, \eta^2 < .01, 95\%CI[-0.40, 0.38]$, on Trust. An exploratory Bayesian ANOVA revealed moderate evidence in favor of the null hypotheses for Disclosure Timing ($BF_{01} = 4.11$) or Strength of Evidence ($BF_{01} = 4.68$). Therefore, we found no difference in the perceived rapport between those to whom evidence was presented late vs early ($M = 3.34, SD = 0.78$ vs. $M = 3.24, SD = 0.99, d = 0.11, 95\%CI[-0.28, 0.50]$) nor weak vs strong evidence ($M = 3.27, SD = 0.92$ vs. $M = 3.30, SD = 0.86, d = -0.03, 95\%CI[-0.42, 0.36]$).

We did, however, observe a significant interaction effect between Disclosure Timing and Evidence Strength, $F(1, 97) = 4.86, p = .03, \eta_p^2 = .05$. Tests of simple effects revealed no statistically significant differences between any of the categories. However, there were near significant differences when comparing Weak to Strong evidence when there was Early Disclosure ($M = 3.03, SD = 0.97$ vs. $M = 3.45, SD = 0.97, d = -0.43, 95\%CI[-0.83, -0.04]$; $F(1,97) = 2.20, p = .09, \text{Mean difference}[95\%CI] = 0.42[-0.73, 0.90]$), which indicated that trust may be lower when Weak versus Strong evidence is revealed early. There was slightly stronger evidence that Trust is higher when weak evidence is released Late vs. Early ($M = 3.51, SD = 0.81$ vs. $M = 3.03, SD = 0.97$; $F(1,97) = 3.87, p = .05, \text{Mean difference}[95\%CI] = 0.48[-0.00, 0.97], d = 0.54, 95\%CI[0.14, 0.93]$). Trust did not differ significantly when Strong evidence was presented Late ($M = 3.16, SD = 0.72$) to any of the other experimental groups.

Overall, we did not find any clear evidence for effects of evidence disclosure timing or evidence strength on trust, but it may be the case that when presenting evidence early, stronger evidence increases trust while weaker evidence decreases it, while the effect is reversed with late disclosure. Weak evidences revealed late is associated with higher trust and strong evidence with lower trust.

Discussion

Key Findings

Overall, we found clear support that late disclosure of evidence produces both more statement-evidence and within-statement inconsistencies. Crucially, we also showed that statement-evidence inconsistencies are especially low when strong evidence is revealed early in the interview. However, neither evidence strength nor disclosure timing had any effect on the amount of novel investigative information revealed. We also found no evidence to suggest that evidence strength or disclosure timing had any negative impact upon the interpersonal relationship between interviewer and interviewee, at least regarding rapport, with only weak indications of any effects on trust where early disclosure of weak evidence may be detrimental.

Statement-Evidence Inconsistencies

As hypothesized, early disclosure of evidence did produce fewer statement-evidence inconsistencies than late disclosure. The argument within the SUE literature to explain this effect is that early disclosure provides suspects with an opportunity to create a plausible account that covers each piece of evidence ((Hartwig et al., 2014; Hartwig et al., 2005; Oleszkiewicz & Watson, 2021). Following Brimbal and Luke (in press), we reasoned that if this is why statement-evidence inconsistencies are lower with early disclosure, then suspects should be especially motivated to make sure their story aligns with stronger over weaker evidence against the suspect. We also found evidence for this hypothesis because statement-evidence inconsistencies were significantly lower when strong evidence was disclosed early compared to weak evidence. This is a clear indicator that participants felt able to contradict or avoid accounting for evidence that they considered weak.

One important observation is that most of the statement evidence inconsistencies, regardless of in which experimental group they occurred, represented omissions rather than direct contradictions of the existing evidence (e.g., “I did not touch anything that does not belong to me”). These findings are in line with assumptions that suspects would primarily adopt an avoidant strategy prior to the disclosure of evidence (Granhag & Hartwig, 2008), and more general findings that liars tend to provide fewer details (DePaulo et al., 2003). However, it was notable that suspects almost never opted to directly lie in our sample. Less than 2% of the statement-evidence inconsistencies we observed contradicted rather than omitted evidence. This is likely reflected in our experimental design which, unlike a genuine suspect interview, did not leave much scope to probe at suspect’s responses, or directly challenge the

responses provided by suspects. It is noticeable that later studies investigating SUE have moved from testing simple crimes to more complex crimes, as well as incorporating more involved interviewing including the gradual disclosure of evidence throughout the interview (Oleszkiewicz & Watson, 2021). We defer to these studies in their ability to fully test the shift from avoidance to escape strategies, but feel it is important to share the methodological note that our paradigm did not elicit many explicit contradictions or denials.

An additional useful insight is that one of the pieces of evidence (being logged onto a library computer) was rarely denied in either early or late disclosure conditions. Based on the research of Hartwig et al. (2005), the suspects were told that they were brought in for questioning because the police know they were ‘in the area’ on the day of the crime. Here we also find support for the findings of Brimbal and Luke (in press) who also found that evidence that varies in terms of reliability (e.g., an eyewitness vs CCTV) mattered more for eliciting inconsistencies than evidence that varied in terms of proximity (e.g., being shown to be in the library vs. at a specific computer). Interestingly however, we found participants frequently denied our second piece of evidence, which also varied in terms of proximity (fingerprint on a table versus fingerprint on a bag). However, it must be said that the difference in proximity between these two elements is very close physically, even if from a legal perspective there is a more meaningful gap between being at the scene of a crime and directly interacting with the victim’s possessions. Perhaps the reason for our eliciting omissions and denials here is that both pieces of evidence are physical, and so are likely to have been interpreted by participants as having high reliability. Consequently, the suspects may have felt a need to address a highly reliable piece of evidence and any variation in proximity was only of secondary importance.

Together with our observations that most suspects chose to omit rather than lie, our consideration of the type of evidence suspects felt a need to deny suggests that not all statement-evidence inconsistencies are equal (Brimbal & Luke, in press). In other words, the nature of inconsistencies may matter as much as the quantity of inconsistencies.

Overall, our findings give good support to the arguments that late disclosure of evidence produces statement-evidence inconsistencies because participants wish to avoid mentioning incriminating details when they are unsure what evidence the police have for them. We also show that one reason the early release of information is problematic is because suspects can provide an account that explains away the evidence against them, and we have shown that participants are especially motivated to do so when the evidence against them is strong. Finally, we have provided further evidence that suggests that the reliability of evidence matters more than how proximally evidence ties a suspect to a crime scene.

Within-Statement Inconsistencies

We also support previous findings that late disclosure of evidence leads to more within-statement inconsistencies – a cue to deception – than early disclosure (Hartwig et al., 2014; Oleszkiewicz & Watson, 2021). In the case of early disclosure, hardly any within-statement inconsistencies were made because participants could construct an account that accounted for all the evidence against them. With late disclosure, more within-statement inconsistencies were made because suspects adjusted their story to account for omissions to their story (or rarely, an outright lie or denial).

Contrary to our predictions, evidence strength had no bearing on the elicitation of within-statement inconsistencies. We expected participants to be more motivated to change their accounts when presented with strong evidence over weak evidence late in the interview. However, it appears motivation to account for inconsistencies with evidence by changing an initial account once revealed are not dependent on evidence strength. Rather the presentation of any evidence, weak or strong, prompts changes in suspects' stories.

Novel Investigative Information

The main goal in any suspect interview is the recovery of important crime-relevant information (Walsh & Bull, 2015). Consequently, analyses of SUE have moved from only comparing how well it performs as a deception detection method toward evaluating whether it can also help to facilitate disclosures from suspects (Luke & Granhag, 2022; May et al., 2017; Tekin et al., 2016). Thus far, laboratory studies have not shown convincingly either way whether SUE does or does not help to facilitate disclosures (Oleszkiewicz & Watson, 2021). In line with the overall findings of the meta-analysis by Oleszkiewicz and Watson (2021), we found no significant effects of either disclosure timing or evidence strength on the number of pieces of novel investigative information suspects shared.

However, we would not yet suggest that SUE is ineffective for eliciting information. The use of SUE as an information elicitation paradigm is still developing, and early indicators suggest that both how evidence is disclosed and how inconsistencies are challenged are likely to be very important in whether evidence disclosure leads to enhanced cooperation. For example, the recent work of Luke and Granhag (2022) suggests that the non-accusatory challenging of inconsistencies can facilitate information disclosure to a greater extent than only generating statement-evidence inconsistencies which are not directly challenged. In other words it is not enough to generate inconsistencies and hope that these will provoke a change in the suspects strategy, the interviewer should take an active role in facilitating this desired

behavioral change (May et al., 2017), and this can be done without resorting to coercive or accusatory methods.

One observation we made is that our suspects spent much time talking around the evidence. That is, many suspects gave elaborate statements regarding their thoughts and feelings during the time of the crime, but without providing any tangible information that could be of use to any investigation. Thus, it appeared that much of the speech used by the participants was not targeted at directly explaining away the evidence, but at contextualizing the crime from the suspects perspective, perhaps to generate empathy or at least understanding from the interviewer.² This observation aligns with an analysis of the behaviors of suspects during investigative interviews by Watson et al. (2021). Briefly, Watson et al. (2021) found that suspects speech aims to do much more than exonerate themselves from blame, or even otherwise directly account for evidence. Much suspect speech could be classified as relational, and about managing the relationship between suspect and interviewer, but also the interviewer's perceptions of the victim or other relevant parties. Such speech may serve to change the perceptions interviewers have about suspects or victims and so could plausibly affect investigative decision making (Schmuck et al., 2021). We feel it is important to note that suspects counter-interrogation strategies may well be much richer and more nuanced than only deciding which pieces of information should or should not be disclosed. Rather, they also try to shape how the information they disclose is understood and framed by interviewers.

Evidence disclosure and the interviewer-interviewee relationship

One concern about the use of SUE is that it might inhibit the quality of the relationship between the interviewer and suspects (Clemens et al., 2020), either by making interviews conform to a rigid structure and so inhibiting the quality of communication or by making participants feel the interviewer has behaved unfairly by withholding evidence. We did not find convincing evidence that these effects occur. Rapport was entirely unaffected by the late disclosure of evidence, and while we did find evidence for an interaction effect of disclosure timing and evidence strength on trust, the strength of that evidence was rather modest.

The null effects on rapport are important, given the centrality of rapport to securing the cooperation of suspects (Alison et al., 2013; Gabbert et al., 2021; Walsh & Bull, 2012).

² A full analysis of the specific influence behaviours used by our participants is beyond the scope of this paper. However, it was the focus of one of the author's MSc theses which used data from this experiment. That thesis is freely accessible here: <http://essay.utwente.nl/89434/>

However, we would not yet conclude that there could not be any effect of evidence disclosure timing or strength on rapport. Our experiment comprised short, low stakes interviews about a simple crime with few pieces of evidence. The participants were also often familiar with the experimenter in this study, if not the interviewer, but rapport held with the experimenter may have transferred to interviewer (Weiher, 2020). It is certainly plausible that any effect on rapport might change under different conditions. For example, it is very likely much more difficult to build and maintain rapport in the first place when stakes are higher and when genuine suspects are used. The simplicity of the interview is also likely to be relevant. A specific concern raised by practitioners within the research of Clemens et al. (2020) was that a strict adherence to SUE might inhibit the natural flow of communication between suspect and interviewer. Any such effects are likely to be small when interviews are brief and cases are simple. Given these limitations, we think it is important for researchers testing SUE to consider the interpersonal dynamics of the interview participants, especially rapport, so that our promising but tentative findings can be confirmed in suspect samples and with regard to more complex criminal cases.

In contrast to rapport, we did find tentative effects of evidence disclosure timing and strength on the trust suspects held toward their interviewers. Specifically, we found some evidence suggesting that trust is higher when weak evidence is presented late compared to early, and weaker evidence that trust is higher when strong evidence is presented early than when weak evidence is presented early.

Common to both tentative effects is that the lowest level of trust was found for the early disclosure of weak evidence. This means that while the early disclosure of weak evidence may generate more statement-evidence inconsistencies than the early disclosure of strong evidence, this strategy may also inhibit the development of a trusting relationship between interviewer and interviewee. This could be problematic given the importance of developing a cooperative relationship for information disclosure.

Still, we observed only weak (and non-significant) positive correlations between rapport and trust with novel investigative information. The effects of these interpersonal variables are likely to be much more important in interviews of greater length, complexity, and seriousness and in situations where suspects are not participants that have been instructed to try to convince the interviewer of their innocence. That is, our participants did not need to be won over by an interviewer developing a relationship with them to have the suspects provide an account. The consequence of this is that the relationship between rapport and trust in the provision of information is likely to be weaker in our study than in a genuine interview.

Conclusion

We have shown that late disclosure of evidence led to more statement-evidence and within-statement inconsistencies than early disclosure supporting prior research that suggest that this enables SUE to act as a method of deception detection. We also considered the role of evidence strength, and showed that early disclosure of strong (versus weak) evidence leads to fewer statement evidence inconsistencies, likely because suspects feel especially motivated to account for strong evidence. Thus, while our research warns against early disclosure in general, we especially warn against the early disclosure of strong evidence when seeking to elicit statement-evidence inconsistencies. However, the early disclosure of weak evidence may have the most deleterious impact on the level of trust a suspect has with an interviewer. This contrasts with prior assumptions that the late disclosure of evidence would be most likely to have a negative impact on the interpersonal relationship with a suspect.

No effects of disclosure timing or evidence strength was found upon rapport or the provision of novel investigative information. Therefore, while we find support for the use of SUE as a deception detection tool, and tentative evidence SUE does not lower the quality of the interviewer-interviewee relationship, we cannot provide evidence that the basic late disclosure version of SUE leads to suspects providing additional information to their interviewers.

References

- Alison, L., Alison, E., Noone, G., Elntib, S., & Christiansen, P. (2013). Why tough tactics fail and rapport gets results: Observing Rapport-Based Interpersonal Techniques (ORBIT) to generate useful information from terrorists. *Psychology, Public Policy, and Law*, 19(4), 411-431. <https://doi.org/10.1037/a0034564>
- Bond, C. F., & DePaulo, B. M. (2006, August 1, 2006). Accuracy of deception judgments. *Personality and Social Psychology Review*, 10(3), 214-234. https://doi.org/10.1207/s15327957pspr1003_2
- Brimbal, L., & Luke, T. J. (in press). Deconstructing the Evidence: The Effects of Reliability and Proximity of Evidence on Suspect Responses and Counter-Interrogation Tactics. *Journal of Applied Research in Memory and Cognition*. <https://doi.org/https://doi.org/10.1016/j.jarmac.2021.10.001>
- Cassell, P. G., & Hayman, B. S. (1995). Police Interrogation in the 1990s: An Empirical Study of the Effects of Miranda. *UCLA L. Rev.*, 43, 839.
- Clemens, F., Knieps, M., & Tekin, S. (2020, 2020/01/01). Untapped Potential? A Survey Study with German Police Officers into Suspect Interviewing Practices and the Strategic Use of Evidence. *Journal of Forensic Psychology Research and Practice*, 20(1), 53-79. <https://doi.org/10.1080/24732850.2019.1684123>
- Collins, R., Lincoln, R., & Frank, M. G. (2002, 2002/06/01). The effect of rapport in forensic interviewing. *Psychiatry, Psychology and Law*, 9(1), 69-78. <https://doi.org/10.1375/pplt.2002.9.1.69>
- 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(2), 189-202. <https://doi.org/10.1002/jip.145>
- DePaulo, B. M., Lindsay, J. J., Malone, B. E., Muhlenbruck, L., Charlton, K., & Cooper, H. (2003). Cues to deception. *Psychological Bulletin*, 129(1), 74-118. <https://doi.org/10.1037/0033-2909.129.1.74>
- Duke, M. C., Wood, J. M., Bollin, B., Scullin, M., & LaBianca, J. (2018). Development of the Rapport Scales for Investigative Interviews and Interrogations (RS3i), Interviewee Version. *Psychology, Public Policy, and Law*, 24(1), 64-79. <https://doi.org/10.1037/law0000147>
- Fawcett, H. E. (2012). *An investigation into deceptive alibi witness testimony*. [Doctoral thesis, Sheffield Hallam University]. Sheffield Hallam University Research Archive. <http://shura.shu.ac.uk/20199/>
- Gabbert, F., Hope, L., Luther, K., Wright, G., Ng, M., & Oxburgh, G. (2021). Exploring the use of rapport in professional information-gathering contexts by systematically mapping the evidence base. *Applied Cognitive Psychology*, 35(2), 329-341. <https://doi.org/https://doi.org/10.1002/acp.3762>

- Granhag, P. A., Clemens, F., & Strömwall, L. A. (2009). The usual and the unusual suspects: level of suspicion and counter-interrogation tactics. *Journal of Investigative Psychology and Offender Profiling*, 6(2), 129-137. <https://doi.org/10.1002/jip.101>
- Granhag, P. A., & Hartwig, M. (2008, 2008/06/01). A new theoretical perspective on deception detection: On the psychology of instrumental mind-reading. *Psychology, Crime & Law*, 14(3), 189-200. <https://doi.org/10.1080/10683160701645181>
- Granhag, P. A., & Hartwig, M. (2015). The strategic use of evidence technique: A conceptual overview. In *Detecting deception: Current challenges and cognitive approaches*. (pp. 231-251). Wiley-Blackwell.
- Griffiths, A., & Rachlew, A. (2018). From interrogation to investigative interviewing: The application of psychology. In *The Psychology of Criminal Investigation* (pp. 154-178). Routledge.
- 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* (pp. 1-36). Academic Press. <https://doi.org/https://doi.org/10.1016/B978-0-12-394433-7.00001-4>
- Hartwig, M., Granhag, P. A., & Strömwall, L. A. (2007, 2007/04/01). Guilty and innocent suspects' strategies during police interrogations. *Psychology, Crime & Law*, 13(2), 213-227. <https://doi.org/10.1080/10683160600750264>
- Hartwig, M., Granhag, P. A., Strömwall, L. A., & Kronkvist, O. (2006, October 01). Strategic use of evidence during police interviews: When training to detect deception works [journal article]. *Law and Human Behavior*, 30(5), 603-619. <https://doi.org/10.1007/s10979-006-9053-9>
- Hartwig, M., Granhag, P. A., Strömwall, L. A., & Vrij, A. (2005, August 01). Detecting deception via strategic disclosure of evidence [journal article]. *Law and Human Behavior*, 29(4), 469-484. <https://doi.org/10.1007/s10979-005-5521-x>
- Hingmann, N. M. (2019). *Examining an adaptation of the Strategic Use of Evidence Framework when interviewing suspects* University of Twente]. <http://purl.utwente.nl/essays/78972>
- Kassin, S. M., & Fong, C. T. (1999, 1999/10/01). "I'm Innocent!": Effects of Training on Judgments of Truth and Deception in the Interrogation Room. *Law and Human Behavior*, 23(5), 499-516. <https://doi.org/10.1023/A:1022330011811>
- Luke, T. J., & Granhag, P. A. (2022). The shift-of-strategy (SoS) approach: using evidence strategically to influence suspects' counter-interrogation strategies. *Psychology, Crime & Law*, 1-26. <https://doi.org/10.1080/1068316X.2022.2030738>
- May, L., Granhag, P. A., & Tekin, S. (2017, 2017-July-17). Interviewing Suspects in Denial: On How Different Evidence Disclosure Modes Affect the Elicitation of New Critical

- Information [Original Research]. *Frontiers in Psychology*, 8. <https://doi.org/10.3389/fpsyg.2017.01154>
- Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust [Article]. *Academy of Management Review*, 20(3), 709-734. <https://doi.org/10.5465/AMR.1995.9508080335>
- Moston, S., Stephenson, G. M., & Williamson, T. M. (1992). The effects of case characteristics on suspect behaviour during police questioning. *The British Journal of Criminology*, 32(1), 23-40. <https://doi.org/10.1093/oxfordjournals.bjc.a048178>
- Oleszkiewicz, S., & Watson, S. J. (2021). A meta-analytic review of the timing for disclosing evidence when interviewing suspects. *Applied Cognitive Psychology*, 35(2), 342-359. <https://doi.org/https://doi.org/10.1002/acp.3767>
- Sakrisvold, M. L., Granhag, P. A., & Mac Giolla, E. (2017, Jan). Partners under Pressure: Examining the Consistency of True and False Alibi Statements. *Behav Sci Law*, 35(1), 75-90. <https://doi.org/10.1002/bsl.2275>
- Schmuck, A., Wüller, C. A., & Watson, S. J. (2021). The effect of denial of the victim arguments within simulated suspect interviews. European Association of Psychology and Law Conference,
- Sellers, S., & Kebbell, M. R. (2009). When should evidence be disclosed in an interview with a suspect? An experiment with mock-suspects. *Journal of Investigative Psychology and Offender Profiling*, 6(2), 151-160. <https://doi.org/https://doi.org/10.1002/jip.95>
- 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(3), 244-252. <https://doi.org/10.1037/lhb0000131>
- Tekin, S., Granhag, P. A., Strömwall, L. A., & Vrij, A. (2016). How to make perpetrators in denial disclose more information about their crimes. *Psychology, Crime & Law*, 22(6), 561-580. <https://doi.org/10.1080/1068316X.2016.1168425>
- Walsh, D., & Bull, R. (2012). Examining rapport in investigative interviews with suspects: Does its building and maintenance work? [journal article]. *Journal of Police and Criminal Psychology*, 27(1), 73-84. <https://doi.org/10.1007/s11896-011-9087-x>
- Walsh, D., & Bull, R. (2015, 2015/08/09). Interviewing suspects: examining the association between skills, questioning, evidence disclosure, and interview outcomes. *Psychology, Crime & Law*, 21(7), 661-680. <https://doi.org/10.1080/1068316X.2015.1028544>
- Walsh, D., Milne, B., & Bull, R. (2016, 2016/06/01). One Way or Another? Criminal Investigators' Beliefs Regarding the Disclosure of Evidence in Interviews with Suspects in England and Wales. *Journal of Police and Criminal Psychology*, 31(2), 127-140. <https://doi.org/10.1007/s11896-015-9174-5>
- Watson, S. J., Luther, K., Taylor, P., & Jackson, J. (2021). *The Influence Strategies of Interviewees Suspected of Controlling or Coercive Behavior*.

Weiher, L. (2020). *The impact of interview style on the development, maintenance, and transfer of rapport* [Lancaster University]. Lancaster.

Wells, G. L., Small, M., Penrod, S., Malpass, R. S., Fulero, S. M., & Brimacombe, C. A. E. (1998, 1998/12/01). Eyewitness Identification Procedures: Recommendations for Lineups and Photospreads. *Law and Human Behavior*, 22(6), 603-647.
<https://doi.org/10.1023/A:1025750605807>