A meta-analytic review of the timing for disclosing evidence when interviewing suspects

Simon Oleszkiewicz1 | Steven J. Watson2

1Department of Criminal Law and Criminology, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands
2Department of Psychology of Conflict, Risk, and Safety, University of Twente, Enschede, The Netherlands

Abstract
This meta-analytic review examines the most fundamental question for disclosing evidence during suspect interviews: What are the effective options for when to disclose the available evidence? We provide an update to Hartwig and colleagues (2014) meta-analysis of the efficacy of the late and early disclosure methods on eliciting statement-evidence inconsistencies from guilty and innocent suspects. We also extend these analyses to include studies comparing gradual disclosure to early and late disclosure when interviewing guilty suspects. Finally, we test whether a gradual disclosure leads to greater provision of novel investigative information when interviewing guilty suspects. Overall, we find that guilty suspects provide more statement-evidence inconsistencies than innocent suspects, and that both a late and gradual disclosure result in more statement-evidence inconsistencies than the early disclosure when interviewing guilty suspects. However, there are indications of small study effects that warrant considerable caution when interpreting the size of some of the identified effects.

1 | INTRODUCTION

A turning point in the research on deception detection came when DePaulo and colleagues (2003) showed that verbal and non-verbal cues to deception are scarce, and that the extant cues to deception are faint and unreliable in the context of passive observation (Bond & DePaulo, 2006). Two years after DePaulo’s meta-analysis, Maria Hartwig et al. (2005) demonstrated that interviewers could take a more active lie-catching role by withholding available evidence from the suspect until after the suspect had provided an initial account. For interviewing guilty suspects, such “late disclosure” of evidence increased the number of statements that contradicted the available evidence, whereas the statements made by innocent suspects would remain similarly consistent with the evidence regardless of whether the evidence was withheld or presented up front (Hartwig et al., 2005). As deception research continued to demonstrate challenges with passive lie-catching (e.g., Hartwig & Bond, 2011, 2014; Luke, 2019; Vrij, 2008) researchers started arguing against the need to continue the traditional studies on deception detection and encouraged the development of methods designed to elicit and enhance diagnostic cues to deceit (Vrij & Granhag, 2012). This led to an increase in studies examining the efficacy of evidence disclosure as a method to detect deception, such as the Strategic Use of Evidence technique (Hartwig et al., 2014) and the Tactical Use of Evidence technique (Dando & Bull, 2011). More recently, however, researchers have started moving away from utilizing evidence as a means to discriminate between guilty and innocent suspects, and instead attempt to improve investigative outcomes by focusing on the behavior of guilty suspects and their release of novel investigative information (Tekin et al., 2015; Tekin et al., 2016; May et al., 2017).

As might be inferred from above, the studies into evidence disclosure are quite diverse, using various manipulations of disclosure...
methods (cf. Hartwig et al., 2005; Dando & Bull, 2011; Granhag, Strömwall, Willén, & Hartwig, 2013; Luke et al., 2014), a number of specific dependent measures (cf. Dando & Bull, 2011; Hartwig et al., 2005; Luke et al., 2015; Tekin et al., 2015), and different experimental paradigms (cf. Dando et al., 2015; Hartwig et al., 2005; Luke et al., 2015; Wachi et al., 2017). Hence, conceptualizing all strategic and tactical nuances of the extant studies would be much too ambitious for the present review. Nevertheless, we will make a first step in bringing this literature together by examining the timing for when to disclose evidence to suspects. To achieve this, we will (i) provide some theoretical background supporting the two most prevalent research programs into evidence disclosure methods, (ii) clarify the main dependent measures that have been used to examine efficacy, and (iii) explore different criminal scenarios that have been modeled in experimental paradigms. We then meta-analyze the studies to examine which evidence disclosure timing is most effective for (a) eliciting cues to deception and (b) gathering novel investigative information.

2 | THE STRATEGIC USE OF EVIDENCE (SUE) FRAMEWORK

The research program into the SUE technique sets out to examine principles for how evidence can be utilized when interviewing suspects of crime. The theoretical framework of the technique is divided into a strategic level and a tactical level (Granhag & Hartwig, 2015). The strategic level refers to more general and case-independent principles (i.e., the suspect's perception, strategies, and responses), whereas the tactical level refers to more specific and case-dependent tactics (i.e., evidence assessment, question tactics, and disclosure tactics). The backbone of this research is the notion that suspects adopt strategies when attempting to deliver a convincing account, and that the strategies adopted by guilty suspects are different to the strategies adopted by innocent suspects (Granhag & Hartwig, 2008, 2015; Vrij & Granhag, 2012). For example, studies have shown that suspects who lie tend to keep their stories simple while attempting to avoid and deny critical details, whereas suspects who tell the truth are prone to tell the truth as it happened in the hopes that their honesty will be vindicating (Granhag et al., 2015; Leins et al., 2013; Strömwall & Willén, 2011). In essence, the SUE framework outlines a number of strategic and tactical principles for how interviewers can utilize evidence in a manner that identifies and challenges possible deceptive attempts while continually facilitating truthful disclosure.

The principles of this SUE framework draws on the psychological theory of self-regulation (Carver & Scheier, 2012) to explain how guilty suspects can adopt “avoid” strategies (e.g., omit details) and “deny” strategies (e.g., modify details) when attempting to deliver a convincing account (Granhag & Hartwig, 2008, 2015). The argument is that guilty suspects who remain uninformed of the evidence against them should make more statements that contradict the evidence compared to guilty suspect that are made aware of the evidence. Moreover, recent advancements to the SUE framework have continued to draw on the theory of self-regulation to suggest that guilty suspects may feel the need to change their initial strategy if they perceive that the strategy has not been working to their favor (Granhag & Hartwig, 2015, Granhag & Luke, 2018). In order to prevent making the same mistakes again and again, guilty suspects can shift to “escape” strategies (e.g., to shut down and stop communicating) or “repair” strategies (e.g., to become forthcoming with information that the interviewer is perceived to already know). The idea is then, that the interviewer can utilize disclosure strategies and tactics (e.g., to continually point out statement-evidence inconsistencies) and therein influence the suspect's perception of how much information the interviewer already holds, which may lead the suspect to become more forthcoming with novel investigative information (Granhag & Luke, 2018).

3 | THE TACTICAL USE OF EVIDENCE (TUE) FRAMEWORK

Research on the TUE technique stresses that an interviewer's objective is to gather information that will assist decision-makers in the criminal justice system (e.g., prosecutors, juries, and judges). That is, rather than seeking specific information that may influence the interviewer to pass judgment on the suspect, the interviewer should take a more neutral role and work to collect quantities of information that facilitate observers' ability to assess the suspect. Similar to the SUE framework, the TUE technique takes into account that guilty suspects tend to provide shorter and less detailed statements than innocent suspects, but argues that this is because liars have to deal with the limitations of working-memory when constructing, verbalizing, and maintaining verbal accounts. Hence, by keeping their story simple, guilty suspects will provide themselves with opportunities (e.g., to add details when they know it will not incriminate them) and delay time (e.g., to reflect upon how to respond). Moreover, research on TUE emphasizes that innocent suspects can be influenced to not share information in the belief that some details might be irrelevant or unimportant to the interviewer. Taking all the above into account, the idea is that “drip-feeding” the evidence in a gradual procedure throughout the interview will allow observers to continually update their assessment of the suspect. The argument is that a gradual disclosure of evidence will illustrate the suspect's veracity at the outset of the interview, because guilty suspects can be assessed as inconsistent early on whereas innocent suspects have an early opportunity to convey their honesty. From then on observers can calibrate their initial assessment along the course of the interview (Dando et al., 2015).

4 | EVIDENCE DISCLOSURE METHODS

To test strategic and tactical principles of evidence disclosure researchers have made use of three broad disclosure methods that are compared with an early disclosure of evidence. For the early disclosure method, the interviewer discloses all the available evidence
at the outset of the interview, before requesting the suspect to account for what had happened. The early disclosure method is an appropriate control as it allows the suspect to take the interviewer’s knowledge into account when constructing a narrative. Consequently, it is argued that early disclosure plays into the hands of guilty suspects, as this provides them the opportunity to adopt verbal strategies that better mirror those of innocent suspects (Hartwig, 2011).

4.1 Late disclosure

The late disclosure method was introduced to exploit the different verbal behaviors that guilty and innocent suspects adopt when attempting to convince the interviewer of their innocence (Hartwig et al., 2005). For the late disclosure method, the interviewer starts the interview by encouraging the suspect to provide an open-ended narration of the event in question. The interviewer then asks evidence-focused questions designed to make the suspect address the available evidence before being made aware of what evidence the interviewer holds (Hartwig et al., 2011; Hartwig et al., 2014; Luke et al., 2014). It was theorized that withholding the evidence from the suspect in such a manner would increase the number of evidence-contradicting statements made by guilty suspects, compared to innocent suspects, when providing their initial accounts. Hence, the (in)consistency between the suspect’s statements and the available evidence should help discriminate the guilty from the innocent suspects.

4.2 Gradual disclosure

Dando and Bull (2011) made the argument that withholding all the evidence from the suspect until the very end of the interview would make interviewers overly fixated on producing statements that contradict the available evidence, and that such fixation might take a toll on the interviewer’s cognitive resources to consider other features that are central to an effective interview (e.g., having an open mindset and gathering information). They further argued that a gradual disclosure of evidence, in which the available pieces of evidence are disclosed one by one in a continuous procedure throughout the interview, would allow the interviewer more flexibility, and thereby result in a larger variety of verbal cues to deception compared to the early and late disclosure methods, resulting in more accurate veracity assessments.

4.3 Incremental disclosure

Tekin et al. (2015) argued that the aim of suspect interviewing is to increase admissions (e.g., new information that links a suspect with the crime) rather than improving veracity assessments or producing confessions. Drawing from research showing that guilty suspects become more forthcoming when they are aware of the evidence against them (Luke et al., 2014), Tekin et al. (2015) proposed that admissions could be increased by challenging evidence-contradicting statements in an incremental fashion. What is meant by incremental is that the interviewer discloses pieces of evidence and challenges contradictory statements in an order informed by evidence strength (e.g., being around the crime scene versus holding the murder weapon) and source reliability (e.g., a witness statement vs. CCTV footage) (cf. Granhag & Luke, 2018; Granhag et al. 2013). The authors argued that such incremental disclosure procedures could be carefully designed to influence suspects to overestimate the interviewer’s knowledge, and thereby influence the suspect to release isolated pieces of information. Very simply explained, if the interviewer would continually request the suspect to address a specific event for which a piece of evidence is available, and then either affirm or correct the suspect by disclosing the piece of evidence, the suspect may eventually perceive the interviewer to be quite knowledgeable about the situation (May et al., 2017; Tekin et al., 2015; Tekin et al., 2016). When that happens, the interviewer can request a statement that would address an event for which no evidence is available to the interviewer, and the suspect may then be more willing to reveal that information in the belief that the interviewer already holds evidence about it.

In addition to the above, the two incremental dimensions of evidence strength and source reliability can also inform on evidence framing (Granhag et al., 2013), which is about slicing one piece of evidence into two or more units. The units can then be disclosed unit-by-unit with increasing specificity (e.g., “we have information that you entered the shop”; “we have CCTV footage showing that you entered the shop”). Hence, slicing pieces of evidence in this manner can be utilized to influence guilty suspects to contradict their own statements (e.g., “I only passed by the shop, but I was never inside”; “Well, maybe I went in to buy a pack of cigarettes”) (Granhag et al., 2013; Granhag et al., 2015; Luke et al., 2013).

5 MEASURES OF EFFECTIVENESS

To support the efficacy of disclosure methods researchers have made use of various dependent measures to demonstrate specific outcomes. The most commonly used dependent measure is the (in)consistency of the suspect’s account, which can be divided into four sub-categories (Vredeveldt, van Koppen, & Granhag, 2014). Specifically, statement-evidence inconsistency refers to the lack of correspondence between a suspect’s account and the available evidence; within-statement inconsistency refers to the lack of correspondence between details provided in the space of a full account; between-statement inconsistency refers to the lack of correspondence between two consecutive accounts; within-group inconsistency refers to the lack of correspondence between the accounts made by different suspects about the same crime.

Several studies have shown that the timing of evidence disclosure can be used to improve the predictive value of inconsistencies as a cue to deception. For example, to improve the predictive value of statement-evidence inconsistencies the interviewer can disclose
evidence late rather than early in the interview (e.g., Clemens et al., 2011; Hartwig et al., 2005); to improve the predictive value of within-statements inconsistencies the interviewer can slice each piece of evidence into multiple units and then disclose the units gradually (cf. Granhag et al., 2013); to improve the predictive value of within-group inconsistencies a cell of suspects can be separated and interviewed individually in consecutive fashion (Granhag et al., 2015). However, for the purpose of examining the basic timing of evidence disclosure, while also considering the scarcity of evidence disclosure studies examining within-statement inconsistencies (Granhag et al., 2013, Granhag et al., 2015; Luke et al., 2013) and within-group inconsistencies (Granhag et al., 2015; Granhag, Mac Giolla, et al., 2013), the present review will focus exclusively on statement-evidence inconsistencies as a cue to deception.

An outcome measure that has become popular in the more recent literature on evidence disclosure is novel investigative information (e.g., May et al., 2017; Tekin et al., 2015; Tekin et al., 2016). This measure incorporates any critical information that was previously

**TABLE 1** An overview of the studies screened for review based on hypothesis number(s) and exclusion criteria (H/E); experimental paradigm (PA); disclosure timing and incremental tactics; number of evidence pieces available to the interviewer (EP); the existence of un-evidenced criminal phases (UEP); and the main dependent variable (DV)

<table>
<thead>
<tr>
<th>Studies</th>
<th>H/E</th>
<th>N</th>
<th>PA</th>
<th>Disclosure</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Timing</td>
</tr>
</tbody>
</table>
| Included in meta-analysis
| Clemens et al. (2010)* | 1–3 | 84 | A  | Early, late | –  | 2  | No  | SEI   |
| Clemens et al. (2011) | 1–3 | 120| A  | Early, late | –  | 3  | No  | SEI   |
| Granhag et al. (2013)* | 1–5 | 195| A  | Early, late, gradual | Slicing | 1  | No  | SEI   |
| Hartwig et al. (2005)* | 1–3 | 58 | A  | Early, late | –  | 3  | No  | SEI   |
| Hartwig et al. (2006)* | 1–3 | 82 | A  | Untrained, trained | –  | 3  | No  | SEI   |
| Jordan et al. (2012)* | 1–3 | 63 | A  | Early, late | –  | 3  | No  | SEI   |
| Luke et al. (2016) | 1–3 | 59 | A  | Untrained, trained | Slicing | 3  | No  | SEI   |
| Granhag et al. (2015)* | 1–5 | 126| A  | Early, late, gradual | Slicing | 1  | No  | SEI   |
| Luke et al. (2013)* | 1–5 | 102| A  | Early, late, gradual | Slicing | 1  | No  | SEI   |
| Sorochinski et al. (2014)* | 1–5 | 86 | A  | Early, late, gradual | No  | 9  | No  | SEI   |
| Srivatsav et al. (2019) | 4   | 140| B  | None x2 early, gradual | Order | 2  | Yes | SEI   |
| Hartwig et al. (2018) E1 | 4, 6 | 61 | B  | Early, gradual | Order & Slicing | 2  | Yes | SEI, New |
| Hartwig et al. (2018) E2 | 4, 6 | 100| B  | Early, gradual | Order & Slicing | 2  | Yes | SEI, New |
| Hingmann | 5   | 50 | B  | Late, gradual | Order | 5  | Yes | SEI, New |
| May et al. (2017) | 6   | 88 | B  | Early, gradual x2 | Order | 2  | Yes | New   |
| Tekin et al. (2015) | 6   | 90 | B  | None, early, gradual | Order | 6  | Yes | New   |
| Tekin et al. (2016) | 6   | 75 | B  | Early, gradual x2 | Order | 2  | Yes | New   |
| Excluded from meta-analysis
| Lingwood and Bull (2013) | MV  | 52 | A  | Early, late, gradual | No  | 2  | No  | SEI   |
| McDougall and Bull (2015) | MV  | 42 | A  | Early, gradual | No  | 7  | No  | SEI   |
| Clemens and Grolig (2019) | NGS | 94 | A  | Early, late | –  | 2  | No  | SEI   |
| Dando and Ormerod (2018)* | NGS | 54 | A  | Late, late | –  | 2  | No  | New   |
| Dando and Bull (2011) | NDV | 150| B  | Early, late, gradual | No  | 5  | Yes | Veracity |
| Dando et al. (2015) | NDV | 151| B  | Early, late, gradual | No  | 5  | Yes | Veracity |
| Wachi et al. (2017) | NDV | 234| A  | None x2, gradual | No  | 2  | No  | Veracity |
| Luke and Granhag (2020) | NC  | 300| B  | None, gradual x2 | Order & Slicing | 3  | Yes | New   |
| Hartwig et al. (2011) | NM  | 96 | A  | None x3 | –  | 3  | No  | SEI   |
| Luke et al. (2014) | NM  | 143| A  | Late x2 (un/informed) | –  | 2  | No  | SEI   |
| Sukumar et al. (2018) | NM  | 118| A  | None x2 | –  | 3  | No  | SEI   |

Note: * included in Hartwig et al. (2014) meta-analysis.
Abbreviations: -, not applicable; MV, missing value; NC, No relevant control condition for inclusion; NDV, No relevant dependent variable for inclusion; NGS, No guilty suspect; NM, No relevant manipulation for inclusion.
unknown to the interviewer and that might provide new leads for further investigation or establish links between a suspect and a crime (Tekin et al., 2015). Additional outcome measures include forthcominingness (Luke et al., 2014) and veracity assessments (Dando et al., 2015; Dando & Bull, 2011; Wachi et al., 2017). However, these two measures come with limitations, such as a lack in conceptual clarity (Luke et al., 2014) and a failure to specify what informs the global assessment (Dando et al., 2015; Dando & Bull, 2011; Wachi et al., 2017), which might help explain why the measures are rarely used as the main outcome variable.

6 | EXPERIMENTAL PARADIGMS AND MAIN FINDINGS

In order to demonstrate the effects of evidence disclosure methods researchers have developed a number of experimental paradigms to model various criminal scenarios (e.g., theft, cheating, terrorism). These paradigms include a number of sample characteristics, such that most studies have been conducted in the EU and the US, use undergraduates and community members as suspects, and researchers as interviewers (Hartwig et al., 2014). However, the studies also have methodological variations, such as the type of crime the participants are simulating and the number of tasks participants have to perform, and also the amount and type of evidence that is available to the interviewer. Through our reading of the methodologies we identified two broader paradigms that encompassed these methodological intricacies.

6.1 | Paradigm A: The available evidence suggests guilt of a simple crime

This paradigm was originated by Hartwig et al. (2005) and was characterized by three features (see Table 1 for the categorization of studies across the two paradigms). That is, (1) the criminal event commonly entailed a single criminal task (e.g., guilty suspects would steal a book from a bookstore) and (2) would also include a similar but non-criminal task as a comparison group (e.g., innocent suspects would find out the price of the same book in the same bookstore). Furthermore, (3) the interviewer would have evidence that accounted for both the guilty and innocent tasks. For example, a witness had seen the suspect in the bookstore and the suspect’s fingerprints had been found on a book that had laid on top of the stolen book. Hence, the available evidence was equally incriminating towards the guilty suspect as it was towards the innocent suspect. As such, Paradigm A attempts to model criminal activities such as simple theft or cheating in which the existing evidence suggest guilt but may also be perfectly explained by an alternative scenario.

Paradigm A was used in all studies included in the first meta-analytic review comparing the late and early disclosure methods. The meta-analysis included a total of 599 participants and provided strong support that: (i) guilty suspects have a considerably stronger tendency than innocent suspects to make statements that contradict evidence and (ii) guilty suspects’ tendency to contradict evidence nearly doubles in magnitude when the available evidence is disclosed late (Hartwig et al., 2014).

6.2 | Paradigm B: The available evidence fails to establish a complex criminal narrative

This paradigm was originated by Dando and Bull (2011) and later refined by Luke et al. (2015) and was characterized by three features. That is, (1) the criminal event commonly entailed multiple tasks (e.g., stealing and distributing equipment), and the tasks were (2) to be performed across a series of phases (e.g., activities performed before, during, and after the crime). The interviewer (3) had evidence that would account for some, but not all, of the tasks and phases that the suspects performed. As such, Paradigm B attempts to model criminal activities that demand more preparation and are more complex in their execution compared to Paradigm A. Paradigm B may therefore be more similar to criminal activities such as kidnapping, robbery, or terrorism, in which there may exist several pieces of evidence across multiple events.

A number of studies have drawn on the features of Paradigm B to demonstrate the efficacy of the gradual disclosure method for enhancing the accuracy of veracity assessments (Dando et al., 2015; Dando & Bull, 2011), and the incremental method for influencing guilty suspects to become more forthcoming (Luke et al., 2015) and admit self-incriminating information about their activities (May et al., 2017; Tekin et al., 2015; Tekin et al., 2016).

7 | THE TIMING FOR DISCLOSING EVIDENCE

To reiterate, experimental studies have used four basic methods for disclosing evidence in the interrogative context. Three of these methods are fairly straightforward as the timing for disclosure is specified by their label: Early disclosure suggests that the interviewer disclose all available evidence at the outset of the interview (e.g., Hartwig et al., 2005). Late disclosure suggests that the interviewer attempts to collect the suspect’s account before disclosing the evidence, or attempts to collect the account without disclosing the evidence at all (e.g., Hartwig et al., 2005). Gradual disclosure suggests that the interviewer disclose one piece of evidence, and requests the suspects to address that piece of evidence, before disclosing another piece of evidence tied with another request, and then continues this procedure throughout the interview (Dando & Bull, 2011). The incremental method, however, is not labeled by its timing but is commonly categorized on a more tactical level. For example, by (i) the slicing of a single piece of evidence into multiple units (Granhag, Strömwall, Willén, & Hartwig, 2013; Luke et al., 2013; Granhag et al., 2015) and (ii) the order for when to disclose what piece of evidence (May et al., 2017; Tekin et al., 2015; Tekin et al., 2016). Importantly, although the incremental method examines tactical aims rather than general timing, the implementation of the tactical aims suggests that the interviewer
discloses the available pieces (or the units of pieces) in a step-by-step procedure as occurs in the gradual disclosure method. Furthermore, regardless of whether the incremental method examines more specific tactical aims, the tactics fall under the more general concept of timing, and it is important to investigate the typical outcome of disclosure timing across specific tactical implementations.

8 | THE PRESENT REVIEW

The present review makes a first attempt to bring together the literature on evidence disclosure methods by examining the general timing for when to disclose evidence to suspects. Our review of the current state of the research hints to an inherent complexity with respect to comparing studies, which is often due to the fact that studies tend to highlight specific findings without reporting on established dependent measures. Nevertheless, our literature review suggests that the timing of evidence disclosure can be meta-analyzed by dividing hypotheses into three categories. (1) Examining how well statement-evidence inconsistencies discriminate between guilty and innocent suspects; (2) establishing effective evidence disclosure timings for eliciting statement-evidence inconsistencies from guilty suspects; and (3) exploring how evidence disclosure timing influences guilty suspect to reveal novel investigative information. Moreover, we will run our classification of experimental paradigms as a moderator to explore how a simpler versus a more complex criminal simulation may affect the reliability of the observed effect sizes.

To examine how well statement-evidence inconsistencies discriminate between guilty and innocent suspects we followed the footsteps of Hartwig et al. (2014). That is, we included studies comparing the late disclosure and early disclosure methods on statement-evidence inconsistencies (e.g., Clemens et al., 2010; Jordan et al., 2012). This resulted in the following hypothesis:

1. Guilty suspects will provide more statement-evidence inconsistencies than innocent suspects. The magnitude of the difference in the number of statement-evidence inconsistencies between guilty and innocent suspects will be larger when evidence is disclosed late than when evidence is disclosed early.

To establish effective evidence disclosure timings we examined the early, late and gradual disclosure methods on eliciting statement-evidence inconsistencies from guilty suspects. Hence, Hypothesis 2 is in line with Hartwig et al.’s (2014) meta-analysis, whereas Hypotheses 3 and 4 advances the prior meta-analysis (e.g., Granhag et al., 2015; Sorochinski et al., 2014).

2. Late disclosure of evidence will elicit more statement-evidence inconsistencies than early disclosure of evidence. The magnitude of the difference in the number of statement-evidence inconsistencies between late and early disclosure will be larger for guilty than for innocent suspects.

3. Gradual disclosure of evidence will elicit more statement-evidence inconsistencies from guilty suspects than an early disclosure of evidence.

4. Late disclosure of evidence will elicit more statement-evidence inconsistencies from guilty suspects than a gradual disclosure of evidence.

Finally, we made one Hypothesis regarding the timing of evidence disclosure on gathering novel investigative information from guilty suspects (e.g., Tekin et al., 2015). It should be noted that we attempted to directly compare all three disclosure timings (early, late, and gradual) on gathering novel investigative information. However, we found that only Hingmann (2019) included late disclosure as an intervention when novel investigative information was reported as a dependent measure.

5. Gradual disclosure of evidence will elicit more novel investigative information from guilty suspects than an early disclosure of evidence.

We test each of these hypotheses via meta-analysis.

9 | METHOD

9.1 | Inclusion criteria

To test hypothesis 1 and 2 we used the same inclusion criteria as was used by Hartwig, Granhag, and Luke et al. (2014). That is, the included studies had to be laboratory experiments that manipulated evidence to be disclosed early and late in an interrogative context. Early disclosure entailed that evidence was revealed to the suspect at the outset of a scripted interview or, in the case of training studies (i.e., when professional interviewers participated as interviewers in the study), disclosed in any method selected by the interviewers included in the group that had not received training in the late disclosure method prior to testing. Late disclosure entailed that evidence was revealed to the suspect after specific questioning in a scripted interview, or in the case of training studies, in any method selected by the interviewers included in the group that had received training in the late disclosure method prior to testing. Furthermore, the studies had to include a manipulation where participants in the “guilty” condition engaged in some form of mock criminal activity (and may also manipulate an “innocent” condition), the interviewees had to be research participants, and the study had to report a quantitative measure of statement-evidence inconsistencies. The same inclusion criteria were used for testing Hypothesis 3 and 4, with the exception of adding the gradual disclosure method in place of either early or late disclosure methods. Gradual disclosure entailed that the suspect was encouraged to address one piece (or units of a piece) of evidence in relation to its disclosure before being requested to address another piece (or unit) in relation to its disclosure. That is, in this review a gradual disclosure incorporates what Dando and Bull (2011) refer to as gradual or tactical disclosure as well as what Granhag et al. (2013) refer to as incremental disclosure. One modification was made for testing Hypothesis 5: the study had to report a quantitative measure of novel investigative information instead of statement-evidence inconsistencies. Otherwise all previous inclusion criteria applied.
9.2 | Literature search and characteristics of the literature

A literature search was carried out in June 2020. The databases PsycINFO and Google Scholar were searched with the following term: "suspect" AND ("interviewing" OR "interrogation") AND “evidence” AND (“strategy” OR “tactical”) AND “investigative.” This resulted in 368 hits on PsycINFO and 103 hits on Google Scholar. The screening of titles and abstracts yielded 25 studies with potential for inclusion. Additionally, the reference lists of reviews in the field of investigative interviewing were searched and known authors in the field were contacted in an attempt to obtain any unpublished manuscripts relevant to this review. This search yielded five unpublished studies with potential for inclusion.

Of the 30 studies for which methodologies were assessed for eligibility, 16 studies (13 published, 3 unpublished) included a total of 17 experimental samples that met the inclusion criteria (see Figure 1). Thirteen samples came from published articles, one sample came from an unpublished PhD thesis (Srivatsav et al., 2019), one sample came from an unpublished Bachelor thesis (Hingmann, 2019), and three samples came from unpublished governmental reports (Dando & Ormerod, 2018; Hartwig et al., 2018). Ten of the 17 samples were eligible for testing Hypothesis 1, 2 and 3 (N = 778). Hence, only two samples (N = 179) were added to the tests originally performed by Hartwig et al. (2014). For the advanced tests, seven of the 17 samples were eligible for testing Hypothesis 4 (N = 309), five of the 17 samples were eligible for testing Hypothesis 5 (N = 236), and five of the 17 samples were eligible for testing Hypothesis 6 (N = 277).

9.3 | Data analysis

All hypotheses were tested via meta-analysis. Random-effects models with the restricted maximum-likelihood estimation method were calculated to account for likely heterogeneity in the data. Hypotheses 1 and 2 require combining multiple estimates from the same study. This introduces dependency into the data which can bias estimates. We account for this by conducting two- and three-level meta-analyses as recommended by Konstantopoulos (2011). Briefly explained, this involves first performing a two-level meta-analysis which only incorporates the sample from which an estimate is drawn as a random effect. This is identical to the normal scenario for random effects meta-analysis. We then perform an additional three level meta-analysis which nests the random effect of sample within an additional random effect of the article from which the sample was drawn. This adjusts the meta-analytic estimates of effect size by accounting for any correlation between estimates attributable to samples sharing a common source.

To best compensate for the small samples typical in the evidence disclosure literature we report standardized mean differences in the form of Hedges’ g for both statement-evidence inconsistencies and novel investigative information (Hedges & Olkin, 1985). Where statement-evidence inconsistencies were reported as statement-evidence consistencies, these results were reverse scored. We report heterogeneity in terms of I². For completeness, we also report Cochran’s Q as a formal test for the presence of heterogeneity. In addition, we test for any effects of proposed moderators using a Wald test (Borenstein & Higgins, 2013).

To assess the risk of publication bias, we check for small study effects via two methods to account for the imprecision of any single measure of small study effects (Lin et al., 2018). First, we use Egger’s regression (Egger et al., 1997) to regress the standard normal deviate
10 | RESULTS

We first describe how the studies identified in our literature review correspond with our conceptualization of the experimental set-ups (i.e., Paradigm A and B). We also describe the tactical variations within our classification of the gradual disclosure method. We then perform meta-analyses in line with our hypothesis tests. Finally, we integrate the findings from our literature review by conducting exploratory analyses testing for any moderation of identified effect sizes based on (i) the experimental paradigm used and (ii) the tactical implementations within the gradual method.

10.1 | Demographics of experimental paradigms

In line with our conceptualization of the literature we initially attempted to discriminate between experimental Paradigms by considering the suspects’ behavior. We coded the number of criminal tasks and phases that the suspect had to perform (one - Paradigm A vs. multiple - Paradigm B) and whether there was an alternative scenario that would explain all the available evidence (Paradigm A). This coding separated 26 of the 29 experimental set-ups. For the three exceptions (Clemens et al., 2011; McDougall & Bull, 2015; Sorochinski et al., 2014) the suspects’ performed (or planned to perform) a simple task over multiple phases, but also had an alternative scenario that could explain all of the available evidence. To more fully crystallize the distinction between Paradigm A and B, we considered the interviewer’s situation, and coded the amount of evidence available to the interviewer and whether there were any un-evidenced phases of the crime. The results showed that it was only ever studies using Paradigm B that included crime phases for which the interviewer had no evidence available. Incorporating these “un-evidenced phases of the crime” into the coding scheme resulted in the three exceptions being coded as Paradigm A (see, Table 1).

For Paradigm A, the interviewer had between one to nine pieces of evidence available (k = 18, M = 2.88, SD = 1.97). However, when excluding the three exception studies (Clemens et al., 2011; McDougall & Bull, 2015; Sorochinski et al., 2014), the interviewer had between one to three pieces of evidence (k = 15, M = 2.20, SD = 0.75). For Paradigm B, the interviewer had between two to six pieces of evidence available (k = 11, M = 3.27, SD = 2.38).

We also note another distinction between studies. Studies on the early and late methods always examine the general timing of disclosure, whereas studies on the gradual method would sometimes also focus on tactical implementation. Hence, when the pieces of evidence where sliced into units we labeled the gradual method as using a “slicing” tactic; when the pieces of evidence were disclosed in a predetermined sequence we labeled the gradual method as using an “order” tactic; when the gradual method was examined on the level of timing we categorized the gradual method as “no tactic” (see, Table 1). In addition, some studies also used a combination of slicing and order tactics. Where this is the case we meta-analyze these studies as using an “order” tactic, because the piece by piece disclosure of evidence is the more critical distinction between tactical implementations for the present meta-analysis.

10.2 | Meta-analyses of hypothesis tests

10.2.1 | Discriminating guilty from innocent suspects (hypothesis 1)

First we test whether guilty or innocent suspects generate more Statement-Evidence Inconsistencies (SEIs). We present the relevant summary statistics for this analysis in Table 2 and we present individual study estimates via a forest plot in Figure 2. Overall, we identify a large effect whereby guilty suspects generate more SEIs than innocent suspects (g [95%CI] = 1.27 [0.92, 1.63]), which remains after adjusting for dependency in the data via a three-level meta-analysis (g[95%CI] = 1.27 [0.86, 1.68]). However, both Egger’s regression (Intercept = 2.37, p = 0.18) and Begg’s rank-correlation (r = 0.35, p = .034) test indicated that small study effects may affect this estimate.

We enhance this analysis by examining evidence disclosure methods (Early vs. Late) as a moderator of the effect of guilt versus innocence on SEIs via subgroup analyses. We also present the summary statistics from the individual subgroups in Table 2. Guilty suspects produced more SEIs regardless of whether evidence was disclosed early (g(95%CI) = 0.86 [0.51, 1.20]) or late (g(95%CI) = 1.72 [1.18, 2.25]). A Wald test shows that the greater number of SEIs generated by guilty suspects compared to innocent suspects when using the late method versus the early method is statistically significant.
TABLE 2  Summary of meta-analyses of the difference in statement-evidence inconsistencies between guilty and innocent suspects

<table>
<thead>
<tr>
<th>Comparison</th>
<th>Disclosure strategy/tactic</th>
<th>Meta-analyses (95%CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>k</td>
<td>g</td>
</tr>
<tr>
<td>Hypothesis 1: Guilt versus Innocence</td>
<td>20</td>
<td>1.27</td>
</tr>
<tr>
<td>Early</td>
<td>10</td>
<td>0.86</td>
</tr>
<tr>
<td>Late</td>
<td>10</td>
<td>1.72</td>
</tr>
</tbody>
</table>

Note: Values of g > 0 indicate more SEIs for guilty compared to innocent suspects.
Note: Statistically significant estimates (p < .05) are highlighted in bold print.

Author(s) and Year                        SMD [95% CI]

Early Disclosure
Luke et al. 2016, 0.36 [-0.41, 1.12]
Clemens et al. 2011, 0.49 [-0.14, 1.12]
Sorochinski et al. 2013, 0.63 [-0.14, 1.41]
Luke et al. 2013, 1.49 [0.60, 2.38]
Jordan et al. 2012, 1.15 [0.41, 1.89]
Granhaag et al. 2015, 0.64 [-0.08, 1.16]
Granhaag et al. 2013, 1.83 [1.24, 2.41]
Clemens et al. 2010, 0.72 [0.29, 0.95]
Hartwig et al. 2006, 1.28 [0.51, 1.95]
Hartwig et al. 2005, 1.49 [-0.28, 2.26]
RE Model for Subgroup (Q = 22.83, df = 9, p = .01; I² = 59.3%)
Late Disclosure
Luke et al. 2016, 0.57 [-0.15, 1.29]
Clemens et al. 2011, 1.95 [1.03, 2.77]
Sorochinski et al. 2013, 1.56 [0.74, 2.38]
Luke et al. 2013, 2.76 [1.64, 3.88]
Jordan et al. 2012, 2.12 [1.22, 3.03]
Granhaag et al. 2015, 0.70 [0.08, 1.33]
Granhaag et al. 2013, 2.85 [1.15, 3.54]
Clemens et al. 2010, 1.23 [0.57, 1.89]
Hartwig et al. 2006, 2.69 [1.85, 3.54]
Hartwig et al. 2005, 1.07 [0.29, 1.85]
RE Model for Subgroup (Q = 43.68, df = 9, p = .00; I² = 78.8%)
RE Model for All Studies (Q = 87.20, df = 19, p = .00; I² = 78.9%)

FIGURE 2  Forest plot summary of individual study effects for statement evidence inconsistencies for guilty versus innocent suspects (hypothesis 1)

TABLE 3  Summary of meta-analyses of the benefits of evidence disclosure strategies in eliciting statement-evidence inconsistencies

<table>
<thead>
<tr>
<th>Comparison</th>
<th>Moderator level</th>
<th>Meta-analyses (95%CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>k</td>
<td>g</td>
</tr>
<tr>
<td>Hypothesis 2: Late versus Early</td>
<td>18</td>
<td>0.38</td>
</tr>
<tr>
<td>Innocent</td>
<td>8</td>
<td>-0.03</td>
</tr>
<tr>
<td>Guilty</td>
<td>10</td>
<td>0.70</td>
</tr>
<tr>
<td>Hypothesis 3*: Gradual versus Early</td>
<td>7</td>
<td>0.81</td>
</tr>
<tr>
<td>Paradigm A</td>
<td>4</td>
<td>1.05</td>
</tr>
<tr>
<td>Paradigm B</td>
<td>3</td>
<td>0.58</td>
</tr>
<tr>
<td>Slicing tactic</td>
<td>3</td>
<td>0.98</td>
</tr>
<tr>
<td>Order tactic</td>
<td>3</td>
<td>0.58</td>
</tr>
<tr>
<td>Hypothesis 4*: Gradual versus Late</td>
<td>5</td>
<td>-0.15</td>
</tr>
</tbody>
</table>

Note: Values of g > 0 indicate more SEIs for Late/Gradual compared to Early disclosure, or for Gradual over Late disclosure.
Note: Statistically significant estimates (p < .05) are highlighted in bold print.
*Please note that Hypotheses 3 and 4 analyze include only guilty suspects within studies.
(b = 0.86, SE = 0.32, p = .008). Indicators of small study effects were larger for the late disclosure samples (Egger’s Intercept = 1.87, p = .062; rτ = .42, p = .108) than the early disclosure samples (Egger’s Intercept = −0.06, p = .956; rτ = .16, p = .601).

All studies within this hypothesis group were Paradigm A and so subgroup analysis comparing Paradigm A to Paradigm B was not possible.

10.2.2 | Examining disclosure timing on statement-evidence inconsistencies

Here we compare different evidence disclosure methods in the extent to which they generate SEIs in guilty suspects. Summary statistics for all relevant meta-analyses are presented in Table 3.

Generating SEIs from guilty suspects via early or late disclosure of evidence (hypothesis 2)

First we test the hypothesis that the late method generates more SEIs than the early method. We perform similar analyses for Hypothesis 2 as for Hypothesis 1, but this time we test whether guilt is a moderator of any observed effects of the late over the early method. Note that it was not possible to calculate effect sizes for all comparisons because some samples had identical numbers of SEIs and 0 standard error for both groups, presumably because all participants in those groups made an identical number of SEIs. This makes it impossible to estimate g for these comparisons. The excluded comparisons were the innocent groups for Clemens et al. (2011) and Granhag et al. (2015). Individual study effects are shown via forest plot in Figure 3.

Overall we do find more SEIs when evidence is disclosed late rather than early (g[95%CI] = 0.38 [0.15, 0.62]). These results were identical for the two- and three-level meta-analysis indicating no additional bias was introduced from including multiple samples from the same article. The effects were non-significant for the innocent subgroup (g[95%CI] = −0.03 [−0.27, 0.20]), with the difference in SEIs originating entirely from the guilty subgroup (g[95%CI] = 0.70 [0.47, 0.93]). Therefore, guilt did act as a moderator of disclosure method on SEI (b = −0.74, SE = 0.17, p < .001). There were no indications of small study effects from Egger’s or Begg’s tests in the main analysis (Egger’s Intercept = 0.03, p = .974; rτ = −0.03, p = .881), innocent subgroup (Egger’s Intercept = 0.16, p = .876; rτ = .07, p = .905), or guilty subgroup (Egger’s Intercept = 0.34, p = .732; rτ = .11, p = .728).

Also as with Hypothesis 1, all studies within this hypothesis group were Paradigm A and so subgroup analysis comparing Paradigm A to Paradigm B was not possible.

Generating SEIs from guilty suspects via early or gradual disclosure of evidence (hypothesis 3)

Next, we test whether the gradual method generates more SEIs than the early method within guilty suspects. Individual study effects are presented in the forest plot in Figure 4. Overall we find that the gradual method does generate more SEIs than the early method (g[95%CI] = 0.81 [0.38, 1.24]). However, both Begg’s and Egger’s tests indicated that small study effects may be present (Egger’s Intercept = 3.57, p < .001; rτ = .71, p = .030).

<table>
<thead>
<tr>
<th>Author(s) and Year</th>
<th>SMD [95% CI]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Innocent Suspects</td>
<td></td>
</tr>
<tr>
<td>Luke et al, 2016</td>
<td>0.02 [-0.74, 0.78]</td>
</tr>
<tr>
<td>Sorochninski et al, 2013</td>
<td>0.23 [-0.50, 0.96]</td>
</tr>
<tr>
<td>Luke et al, 2013</td>
<td>0.28 [-0.53, 1.08]</td>
</tr>
<tr>
<td>Jordan et al, 2012</td>
<td>-0.55 [-1.32, 0.23]</td>
</tr>
<tr>
<td>Granhag et al, 2013</td>
<td>-0.06 [-0.55, 0.43]</td>
</tr>
<tr>
<td>Clemens et al, 2010</td>
<td>-0.32 [-0.93, 0.29]</td>
</tr>
<tr>
<td>Hartwig et al, 2006</td>
<td>0.22 [-0.40, 0.93]</td>
</tr>
<tr>
<td>Hartwig et al, 2005</td>
<td>-0.05 [-0.63, 0.73]</td>
</tr>
<tr>
<td>RE Model for Subgroup (Q = 4.31, df = 7, p = 0.74; I² = 0.0%)</td>
<td>-0.03 [-0.27, 0.20]</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Guilty Suspects</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Luke et al, 2016</td>
<td>0.17 [-0.64, 0.88]</td>
</tr>
<tr>
<td>Clemens et al, 2011</td>
<td>1.45 [0.75, 2.14]</td>
</tr>
<tr>
<td>Sorochninski et al, 2013</td>
<td>0.66 [-0.10, 1.42]</td>
</tr>
<tr>
<td>Luke et al, 2013</td>
<td>0.51 [-0.28, 1.31]</td>
</tr>
<tr>
<td>Jordan et al, 2012</td>
<td>1.34 [0.62, 2.06]</td>
</tr>
<tr>
<td>Granhag et al, 2015</td>
<td>0.54 [-0.08, 1.15]</td>
</tr>
<tr>
<td>Granhag et al, 2013</td>
<td>0.63 [0.13, 1.13]</td>
</tr>
<tr>
<td>Clemens et al, 2010</td>
<td>0.46 [-0.15, 1.08]</td>
</tr>
<tr>
<td>Hartwig et al, 2006</td>
<td>0.88 [0.24, 1.53]</td>
</tr>
<tr>
<td>Hartwig et al, 2005</td>
<td>0.48 [-0.23, 1.20]</td>
</tr>
<tr>
<td>RE Model for Subgroup (Q = 11.39, df = 9, p = 0.25; I² = 16.1%)</td>
<td>0.70 [0.47, 0.93]</td>
</tr>
<tr>
<td>RE Model for All Studies (Q = 36.63, df = 17, p = 0.00; I² = 53.9%)</td>
<td>0.38 [0.15, 0.62]</td>
</tr>
</tbody>
</table>

**FIGURE 3** Forest plot summary of individual study effects for statement evidence inconsistencies for late versus early disclosure of evidence (hypothesis 2)
We also perform subgroup analyses to determine whether there was any effect of Paradigm (A vs. B) or of tactical implementation (Slicing vs. Order). However, we note that all studies that used Paradigm B used order tactics, whereas all studies that used Paradigm A used slicing tactics (with the exception of Sorochinski et al., 2014 who did not examine tactical implementation). Therefore it is very difficult to disentangle any separate effects of experimental paradigm from tactical implementation. Nonetheless, we find a statistically significant advantage for the gradual method over the early method using Paradigm A ($g$[95%CI] = 1.05 [0.25, 1.85]), with a non-significant effect of similar magnitude for studies using evidence slicing ($g$[95%CI] = 0.98 [−0.10, 2.06]). However, both analyses have strong evidence of small study effects through Egger’s regression (Paradigm A: Egger’s Intercept = 4.01, $p < .001$; Slicing: $I^2$ = 81%; Slicing: $r_{τ}$ = 1, $p = .083$). The Begg’s regressions are not statistically significant, but this can be attributed to statistical power and imprecision from a small number of estimates because in both cases less precise studies were uniformly associated with larger effect sizes (Paradigm A: $r_i$ = 1, $p = .083$; Slicing: $r_i$ = 1, $p = .333$). This high heterogeneity and evidence of small study effects contributes to the effect size estimates having wide confidence intervals and the difference in statistical significance observed between the Paradigm A and Slicing studies, despite these comparisons being largely comprised of the same samples. Studies using Paradigm B/Order tactics showed an advantage for the gradual method over the late method ($g$[95%CI] = 0.58 [0.25, 0.91]) without indications of small study effects ($g$[95%CI] = 0.14, $p = .890$; $r_i$ = .33, $p = 1$) or heterogeneity (See Table 3).

A direct comparison of Paradigm A to Paradigm B did not identify any moderation effect via a Wald test; $b = −0.47$, SE = 0.44, $p = .290$. A comparison of slicing versus order tactics also failed to identify any moderation effect ($b = −0.40$, SE = 0.58, $p = .487$).

Generating SEIs from guilty suspects via late or gradual disclosure of evidence (hypothesis 4)

In contrast to Hypothesis 3, we did not identify any superiority of the gradual method over the late method in eliciting SEIs from guilty suspects ($g$[95%CI] = −0.15 [−0.55, 0.25]). We present individual study effects in Figure 5. There were no indications of small study effects for this analysis (Egger’s Intercept = 1.17, $p = .243$; Slicing: $r_{τ}$ = 1, $p = .817$). We do not perform subgroup analyses on this group of studies because only one study used Paradigm B (Hingmann, 2019), and all studies used slicing tactics except for Sorochinski (2014) who did not use tactics, and Hingmann (2019) who used order tactics.

Luke et al. (2013) also present two different gradual methods which are both compared to the same control group. It is possible to construct a meta-analysis which includes both samples, however these results are only meaningful for comparing multiple treatments. Since Luke et al. (2013) are the only study that present multiple treatments we choose to test only the gradual approach within this study which employs a 2-step procedure (a piece of evidence is sliced into two units) rather than a 4-step procedure (a piece of evidence is sliced into four units). We make this choice because this is the group that is most consistent with other studies. Nonetheless for the sake of thoroughness and transparency we present the code to perform the analysis that includes both studies in the online supplement. In line with our main analysis, neither of the two treatment groups indicate

<table>
<thead>
<tr>
<th>Author(s) and Year</th>
<th>SMD [95% CI]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Paradigm A</td>
<td></td>
</tr>
<tr>
<td>Sorochinski et al., 2013</td>
<td>1.33 [0.51, 2.15]</td>
</tr>
<tr>
<td>Luke et al., 2013</td>
<td>2.00 [1.06, 2.94]</td>
</tr>
<tr>
<td>Granhag et al., 2015</td>
<td>1.05 [0.41, 1.70]</td>
</tr>
<tr>
<td>Granhag et al., 2013</td>
<td>0.07 [-0.42, 0.56]</td>
</tr>
<tr>
<td>RE Model for Subgroup (Q = 16.99, df = 3, $p = 0.00$; $I^2$ = 81.0%)</td>
<td>1.05 [0.25, 1.85]</td>
</tr>
<tr>
<td>Paradigm B</td>
<td></td>
</tr>
<tr>
<td>Srivastav et al., 2020</td>
<td>0.58 [0.10, 1.06]</td>
</tr>
<tr>
<td>Hartwig et al., 2018</td>
<td>0.72 [0.04, 1.40]</td>
</tr>
<tr>
<td>Hartwig et al., 2018</td>
<td>0.47 [-0.14, 1.08]</td>
</tr>
<tr>
<td>RE Model for Subgroup (Q = 0.28, df = 2, $p = 0.87$; $I^2$ = 0.0%)</td>
<td>0.58 [0.25, 0.91]</td>
</tr>
<tr>
<td>RE Model for All Studies (Q = 17.88, df = 6, $p = 0.01$; $I^2$ = 68.4%)</td>
<td>0.81 [0.38, 1.24]</td>
</tr>
</tbody>
</table>

**FIGURE 4** Forest plot summary of individual study effects for statement evidence inconsistencies for gradual versus early disclosure of evidence (hypothesis 3)
significant effects (Two-step: $g[95\%\text{CI}] = -0.20$ [−0.47, 0.06]; four-step: $g[95\%\text{CI}] = 0.47$ [−0.21, 1.14]). Similarly, including Luke et al. (2013) sample that experienced the 4-step procedure in place of the 2-step sample has no meaningful effect on the results presented in Table 3 ($g[95\%\text{CI}] = -0.04$ [−0.50, 0.41], SE = 0.23, $p = .850$ Q (4) = 11.47, $p = .022$, $\tau = 0.42$, $I^2 = 65.16$).

### 10.2.3 Influencing guilty suspects to reveal novel investigative information (hypothesis 5)

Finally, we tested whether use of the gradual method lead to guilty suspects revealing more novel investigative information than those exposed to the early method (Hypothesis 5). All studies in

---

<table>
<thead>
<tr>
<th>Author(s) and Year, Paradigm</th>
<th>SMD [95% CI]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Granhag et al, 2013, Paradigm A</td>
<td>-0.52 [-1.02, -0.02]</td>
</tr>
<tr>
<td>Granhag et al, 2015, Paradigm A</td>
<td>0.53 [-0.08, 1.15]</td>
</tr>
<tr>
<td>Luke et al, 2013, Paradigm A</td>
<td>-0.12 [-0.89, 0.65]</td>
</tr>
<tr>
<td>Sorochinski et al, 2013, Paradigm A</td>
<td>0.01 [-0.70, 0.73]</td>
</tr>
<tr>
<td>Hingmann, 2019, Paradigm B</td>
<td>-0.50 [-0.99, -0.00]</td>
</tr>
</tbody>
</table>

**FIGURE 5** Forest plot summary of individual study effects for statement evidence inconsistencies for gradual versus late disclosure of evidence (hypothesis 4)

<table>
<thead>
<tr>
<th>Author(s) and Year, Paradigm</th>
<th>SMD [95% CI]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hartwig et al. (Expt. 1), 2018, Paradigm B</td>
<td>0.05 [-0.55, 0.66]</td>
</tr>
<tr>
<td>Hartwig et al. (Expt. 2), 2018, Paradigm B</td>
<td>-0.44 [-1.11, 0.22]</td>
</tr>
<tr>
<td>May et al., 2017, Paradigm B</td>
<td>0.65 [0.12, 1.18]</td>
</tr>
<tr>
<td>Tekin et al., 2015, Paradigm B</td>
<td>0.46 [-0.06, 0.97]</td>
</tr>
<tr>
<td>Tekin et al., 2016, Paradigm B</td>
<td>0.95 [0.37, 1.54]</td>
</tr>
</tbody>
</table>

**FIGURE 6** Forest plot summary of individual study effects for novel investigative information for gradual versus early disclosure of evidence (hypothesis 5)
this sample used Paradigm B and order tactics so we do not perform any subgroup analyses. We illustrate individual study effects in the forest plot in Figure 6. We do not find any advantage of the gradual method over the early method in eliciting novel information ($g[95\% CI] = 0.35 [-0.10, 0.81], SE = 0.23, p = .125$). Neither Egger’s (Intercept = −1.65, $p = .100$) nor Begg’s test ($r = -0.40, p = .483$) indicated small study effects.

However, we note that three studies (both experiments within Hartwig et al., 2018; Tekin et al., 2016; May et al., 2017) presented a version of the gradual method in which the suspects’ statements were immediately qualified by the interviewer. That is, if suspects made a statement that contradicted known evidence, the interviewer would explain that the suspect’s statement did not match the available evidence. Therefore we also performed a separate meta-analysis of these four samples, and we present individual study effects in Figure 7. Again we also do not identify an advantage of the gradual method versus the early method ($g[95\% CI] = 0.28 [-0.10, 0.66], SE = 0.19, p = .154$). There were also no indicators of small study effects from Egger’s regression (Intercept = 0.19, $p = .848$), Begg’s test ($r = 0, p = .1$).

11 | DISCUSSION

This study set out to collect the existing experimental studies on evidence disclosure methods and examine the options for when to disclose evidence to suspects; early, late, or gradually. The present meta-analytic review supports previous findings that statement-evidence inconsistencies are a reliable cue to deception and that the late disclosure of evidence is effective for distinguishing guilty from innocent suspects (Hartwig et al., 2014). When examining recent advancements in the experimental literature, the present review suggests that the tendency for guilty suspects to make evidence-contradictory statements amplifies to a similar extent when evidence is disclosed gradually as when it is disclosed late when compared to the early disclosure of evidence. However, we failed to establish support for the prediction that guilty suspects would reveal more novel investigative information when evidence is disclosed gradually rather than early.

11.1 | Main findings

11.1.1 | Statement-evidence inconsistencies as a cue to deception

This review provides strong support for the claim that guilty suspects makes more evidence-contradicting statements than innocent suspects, and especially so when facing a late disclosure of evidence method. That is, guilty suspects are more likely to withhold and modify details pertaining to their prior activities than innocent suspects, and the discriminant value of this tendency magnifies considerably when the available evidence is not immediately revealed to suspects. This marks statement-evidence inconsistencies as a reliable cue for distinguishing guilty from innocent suspects.
11.1.2 Establishing effective timings for disclosing evidence

We find that the late disclosure of evidence leads to more statement-evidence inconsistencies than early disclosure of evidence. However, and importantly, we observed no tendency for innocent suspects to become less consistent with the evidence when they are uninformed about the evidence speaking against them. That is, innocent suspects seem to remain similarly consistent with the evidence regardless of whether they face a late or an early disclosure method. This suggests strong evidence that a late disclosure of evidence is effective in magnifying reliable cues to deception from guilty suspects, as late disclosure appears to have a rather neutral impact on innocent suspects in this regard. On a more cautious note, however, there are reasons to believe this estimated effect size may be exaggerated. At face value the late disclosure method seems to influence guilty suspect to make much more contradictory statements than innocent suspects ($g = 1.72$). However, this increase is less impressive when comparing the early and late disclosure methods across guilty suspects ($g = 0.70$). In fact, this latter comparison suggests that the advantage of the late over the early disclosure method relies heavily on the innocent suspects' level of cooperation (e.g., wants to share information and/or understands that all details are relevant) when facing the late disclosure method. Moreover, the estimate showing the more substantial advantage for late disclosure in distinguishing between guilty and innocent suspects (i.e., $g = 1.72$) might also be exaggerated by small study effects. Hence, more research is needed into the late disclosure method.

Similar to the late disclosure of evidence, we also find uncertainty associated with the effects of gradual disclosure leading to more statement-evidence inconsistencies from guilty suspects than does the early disclosure. Importantly, here the uncertainty is accompanied with a strong warning as the larger effects are heavily associated with smaller studies. Hence, the positive support for the efficacy of the gradual disclosure method has to be considered tentative for now.

Relatedly, for the purposes of the present meta-analytic review we categorized all timing related tactical implementations as a gradual timing of disclosure. This categorization is appropriate for comparing the general effect of gradual disclosure and early disclosure on eliciting statement-evidence inconsistencies. However, some tactics are designed to achieve specific purposes that might work against the typical effects of their general timing. For example, the tactic of evidence framing involves presenting each piece of evidence as sliced into units and disclosed with increasing specificity (cf. Granhag et al., 2013). Such evidence slicing is designed primarily to influence guilty suspects to contradict their own statements (i.e., within-statement inconsistencies) rather than to contradict the available evidence (i.e., statement-evidence inconsistencies). This may explain why we did not identify a significant effect of the slicing tactic on statement-evidence inconsistencies. Unfortunately, however, the tests of the slicing tactic had strong indications of small study effects (Granhag, Strömwall, Willén, & Hartwig, 2013; Granhag et al., 2015; Luke et al., 2013), meaning that we cannot be sure whether there is any effect of slicing on statement-evidence inconsistencies.

More generally, this review shows a clear need for further testing of the gradual disclosure method. The number of studies investigating the gradual method are few, particularly when separated into different experimental paradigms and tactical implementations. This means that estimates of effect size and of small study effects are necessarily imprecise. On a more positive note, we find no difference in the elicitation of statement-evidence inconsistencies between the late and gradual disclosure methods with guilty suspects. That is, if gradual disclosure is similarly effective in eliciting cues to deception as late disclosure (which shows reliable effects when compared with early disclosure), gradual disclosure is likely to be superior to early disclosure in eliciting statement-evidence inconsistencies. Hence, we again recommend additional studies with larger samples so that we can be confident in any benefits of the gradual disclosure method.

11.1.3 Influencing guilty suspects to release novel investigative information

Researchers have argued that the benefits of a gradual disclosure method are likely to go beyond eliciting statement-evidence inconsistencies (Tekin et al., 2015; Tekin et al., 2016; May et al., 2017; Granhag & Luke, 2018, see also Dando & Bull, 2011; Dando et al., 2015). However, the available studies do not support that a mere gradual method, in which evidence is disclosed piece-by-piece, would influence guilty suspects to reveal new information. We would nevertheless like to point out that while the null-finding studies did not indicate small study effects for this test, there were only five experiments testing an incremental order within the gradual disclosure method. Therefore, in the same manner as we apply caution to results where small studies may have exaggerated positive effects, for this test we highlight the possibility that negative results (Hartwig et al., 2018) may have reduced any identified positive effects. Again, we recommend additional studies and the use of larger samples so that there can be greater confidence in the effect size estimates generated.

It should however be noted that the present review tested the general effect of timing on new information, whereas the conceptual argument for influencing guilty suspects to release novel information is made on a tactical level (cf. Granhag & Luke, 2018). The idea is that the interviewer can map out a gradual order for requesting the suspect to address events for which pieces of evidence are already available, and then continually affirm or correct the suspect with the corresponding piece of evidence. By doing so the suspect might come to overestimate how much information the interviewer already holds, leading the suspect to become more forthcoming with novel investigative information. We encourage more studies into tactical arguments like these.

11.2 Experimental paradigms

In the context of this review we attempted to classify the existing experimental paradigms used for examining evidence disclosure
methods. Across the 30 studies assessed we categorized experimen-
tal set-ups based on the number of criminal phases that the suspect
performed into Paradigms A and B. Paradigm A modeled more sim-
ple criminal behavior (e.g., theft or cheating) and the interviewer
would commonly have between one to three pieces of evidence
available. The available evidence would clearly suggest that the sus-
pect was guilty of the crime, but the evidence could also be perfectly
explained within the frame of an alternative scenario. Paradigm B
modeled a more complex criminal behavior (e.g., robbery, terrorism)
in which the suspect would have to perform multiple consecutive
phases to accomplish their criminal task. Here, the interviewer
would commonly have between two to five pieces of evidence avail-
able and the evidence would only account for some, but not all, of
the tasks and phases that the suspect had performed. That is, the
evidence would motivate to interview a specific person as a suspect,
but the evidence would not be strong enough to establish criminal
narrative in court.

We made two main observations across the two experimental
paradigms. First, Paradigm A has predominantly been used for exam-
ining the late disclosure of evidence method on eliciting statement-
evidence inconsistencies (Hartwig et al., 2014), whereas Paradigm B
emerged more in the development of the gradual disclosure method
(e.g., Dando & Bull, 2011; Sorochinski et al., 2014; McDougall &
Bull, 2015). This makes sense as the set-ups within Paradigm B have
progressed towards hosting the conditions necessary to examine
guilty suspects’ release of novel investigative information (May
et al., 2017; Tekin et al., 2015; Tekin et al., 2016). Second, we
observed larger effect sizes and between study heterogeneity for
statement-evidence inconsistencies when the available evidence
suggested guilt of a simple crime (Paradigm A) and smaller, but more
consistent effect sizes when the available evidence failed to establish
a criminal narrative (Paradigm B). This observation seems somewhat
paradoxical, as fewer pieces of evidence available to the interviewer
(Paradigm A) should result in fewer opportunities to elicit statement-
evidence inconsistencies. However, this observation may well be
explained by conflating factors, such as the strength and reliability of
the available evidence. That is, in Paradigm A the evidence would
commonly suggest guilt of a single criminal task. For example, the sus-
pect’s fingerprints had been found on a book that was placed on top
of the stolen wallet. Hence, guilty suspects might perceive it impor-
tant to leave behaviors associated with the criminal task unaddressed
to the longest extent possible. For Paradigm B, however, the evidence
would commonly suggest the suspect had engaged in suspicious
behavior across a number of phases, but the actual crime would be
unevidenced. For example, a witness had seen the suspect receive an
object from a known criminal, and another witness had videorecorded
the suspect taking out a cloth bag from a locker. However, there was
no evidence regarding the content of the bag (a bomb) or where it
had been delivered. Hence, guilty suspects in such scenarios might
similarly want to leave the criminal task unaddressed for the longest
extent possible, but might perceive little risk with immediately
addressing other behavior that could be considered suspicious. Con-
sequently, the number of statement-inconsistencies might be
relatively fewer in Paradigm B due to the strength and reliability of
the available evidence rather than the amount of evidence available.

Due to the perfect overlap between the two experimental para-
digms (A and B) and the two gradual disclosure tactics (Slicing &
Order) it is difficult to separate the effects of the tactics from the
impact of their respective paradigm. That is, we cannot be sure if the
observed effect of a tactical implementation is due to the tactic itself,
or if it is due to the conditions provided within that experimental par-
digm. A final concern is that there were indications of small study
effects for studies using Paradigm A, but not for those using Paradigm
B. This is the likely source of at least some of the additional heteroge-
eity in studies using Paradigm A and has reduced the precision of the
effect size for this paradigm. In sum, more research is needed both
within and between different experiments paradigms.

11.3 Future considerations

Although this review continues to support the utility of being strategic
with evidence when interviewing suspects, there is a clear need to
continue researching this topic. Below we highlight two areas for
consideration.

Most importantly, this review raises concern about how reliably the
late and gradual disclosure methods elicit statement-evidence inconsis-
tencies. In addition to our repeated criticism of small study samples we
would also like to stress that only one study has compared the early
versus late disclosure methods on statement-evidence inconsistencies
over the six last years. This is rather alarming as there still exists uncer-
tainty regarding these effects, and specifically so across different mock-
criminal scenarios. We thus strongly encourage the continued testing of
the basic timings for disclosing evidence in the interrogative context,
and we are particularly keen that future studies (a) include much larger
samples for experimental manipulations and (b) report on the
established dependent measures. Our first point is especially important,
because our review suggests it is likely that the late or gradual disclo-
sure confers a benefit over early disclosure of evidence. What remains
unclear is how large that benefit is, which is now the critical question to
address before we can confidently recommend the late and gradual dis-
closure methods for practice. Moreover, we noted that not less than six
of the 28 studies (20%) were excluded from the present meta-analytic
review solely based on a failure to report on statement-evidence inconsis-
tencies or novel investigative information. These two measures are
critical for relating and comparing disclosure methods across specific
strategic and tactical implementations. Hence, we strongly encourage
future studies to report on these dependent measures also when setting
out to examine implementations tailored to other outcomes.

In addition, concerns have been raised regarding how a fixation on
eliciting statement-evidence inconsistencies may affect the overall
adherence to an ethical interview standard (Dando et al., 2015; Dando &
Bull, 2011). This is concerning partly because the diagnostic effective-
ness of the Strategic Use of Evidence (SUE) framework is supported by
the idea that innocent suspects will be quite forthcoming with informa-
tion, which suggests their statements should be relatively consistent
with the evidence. However, the argument that innocent suspects are relatively forthcoming is mainly based on the behavior of student populations, and there are reasons to believe that other populations may be less willing to, and/or not understand that they should, provide information when interrogated as a suspect but are in fact innocent of the specific crime (Gran Hag, Clemens, & Strömwall, 2009; Clemens & Grolig, 2019). To better understand the possible ethical complications related to strategic evidence disclosure there is a need for more specific examinations into different populations of both innocent and guilty suspects (as guilty suspects are also likely to vary in behavior). A way forward may be to draw inspiration from recent developments in the literature in which different types of suspects are exposed to evidence disclosure methods (Clemens & Grolig, 2019). For example, by manipulating innocent and guilty suspects’ affective relationship with the interviewer, researchers may better be able to model behaviors of other populations (see, Oleszkiewicz & Granhag, 2019).

12 | CONCLUSIONS

The present study meta-analyzed the effective options for when to disclose evidence when interviewing suspects. The late and early disclosure methods was compared on eliciting statement-evidence inconsistencies from guilty and innocent suspects. We also extended these analyses to compare the gradual disclosure to early and late disclosure when interviewing guilty suspects. Finally, we tested whether a gradual disclosure leads to greater provision of novel investigative information when interviewing guilty suspects.

Overall, we found that guilty suspects make statements that contradict the available evidence than innocent suspects, and that both a late and gradual disclosure increases the number of evidence-contradicting statements made by guilty suspects compared to an early disclosure. Although much work remains in order to establish the reliability of these effects, the present review clearly supports that it is important that suspects remain uninformed of the evidence against them, and that it may matter less whether the evidence is undisclosed until the very end of the interview or if the evidence is revealed piece-by-piece throughout the interview.

ACKNOWLEDGMENTS

This work was funded by the High-Value Detainee Interrogation Group Contract DJF-15-1200-V-0010404 awarded to Iowa State University. Statements of fact, opinion, and analysis in the paper are those of the authors and do not reflect the official policy of the High-Value Detainee Interrogation Group or the U.S. Government.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available in the Open Science Framework at https://osf.io/fjk49/. We also supply the R code used for all analyses in the same location.

REFERENCES


APPENDIX CODE

dat <- data. Frame(study = c(1, 2, 3, 4, 5), trt = c(1, 1, 2, 1, 1),
Author = c("Granhag et al," "Granhag et al,
  "Luke et al (2-Step)," "Luke et al (4-Step),"
  "Sorochinski et al," "Hingmann"),
m1i = c(5.48, 1.52, 2.67, 3.38, 20.27, 3.83),
m2i = c(6.15, 0.67, 2.85, 2.85, 20.20, 4.27),
sdpi = c(1.2750098, 1.5669716, 1.4268847, 1.0462982,
  5.1727942, 0.8753499),
n1i = c(32, 21, 13, 14, 15, 24), n2i = c(32, 21, 13, 13, 15, 50)).

dat$Ni <- unlist(lapply(split(dat, dat$study), function(x) rep(sum(x$n1i) + x$n2i[1], each = nrow(x)))).

dat$yi <- with(dat, (m1i-m2i)/sdpi).
dat$vi <- with(dat, 1/n1i + 1/n2i + yi^2/(2*Ni)).
dat
calc.v <- function(x) {
  v <- matrix(1/x$n2i[1] + outer(x$yi, x$yi, "*")/(2*x$Ni[1]),
  nrow = nrow(x), ncol = nrow(x)).
  diag(v) <- x$vi.
  v
}
V <- bldiag(lapply(split(dat, dat$study), calc.v)).
V
H5 <- rma.mv(yi, V, mods = ~ factor(trt) - 1, data = dat).
H5