The prevalence of false confessions in experimental laboratory simulations: A meta-analysis

Joshua M. Stewart | William Douglas Woody | Steven Pulos

1 | INTRODUCTION

False confessions are antithetical to modern systems of justice. They circumvent justice and harm individuals. Harm can include loss of life, freedom, property, and employment. Unfortunately, false confessions do occur, and with disastrous results, leading not only to extremely high primary costs (i.e., to the falsely convicted defendant and to the community which carries the risks of a free perpetrator for whom police no longer search), but also to high secondary costs, both to the legal system and to society at large (e.g., loss of credibility for the police, loss of public confidence in the criminal justice system, and loss of funds used to prosecute an innocent defendant; for reviews see, e.g., Kassin et al., 2010; Kassin & Gudjonsson, 2004; Woody, Forrest, & Stewart, 2011).

Estimates vary regarding the prevalence of false confessions during police interrogation (see Kassin, 1997). West and Meterko (2016) report that 12% of the first 325 individuals DNA-exonerated by the Innocence Project had falsely confessed. Garrett’s (2008) review of the first 250 individuals exonerated by the Innocence Project revealed that 16% included a false confession. Similarly, out of 2,128 exonerated individuals listed in the National Registry of Exonerations (2017), 256 individuals (12%) falsely confessed. Other estimates of false confession rates suggest that the problem remains widespread. In a comprehensive study of police detectives, interrogators estimated that 4.78% of innocent suspects provide a partial admission or a complete confession (Kassin et al., 2007). Additionally, Gudjonsson, Sigurdsson, Sigfusdottir, and Young (2012) surveyed 11,388 Icelandic youths and young adults (95% aged 16–24 years). The researchers found that 12.4% of those who had been interrogated by police reported that they had falsely confessed; rates of 7–12% have been reported in similar surveys (e.g., Gudjonsson, Sigurdsson, Asgeirsdottir, & Sigfusdottir, 2006; Steingrimsdottir, Hreinsdottir, Gudjonsson, Sigurdsson, & Neilson, 2007).

Scholars have used at least three different methods to investigate false confessions. First, case studies of individual false confessors remain invaluable. For example, Garrett (2010) examined false confession cases and how such confessions were litigated at trial. Such studies are rich in detail and ecologically valid. Unfortunately, their very
richness and complexity limit generalizability and prevent the identification of specific factors responsible for false confessions.

Second, correlational studies seek to identify personality characteristics of those individuals who engage in false confessions (see Gudjonsson, 1989). For example, Gudjonsson, Sigurdsson, Einarsson, Bragason, and Newton (2010) administered personality tests to male prisoners, including some who reported a history of giving false confessions to police, and the researchers found that inattention and hyperactivity symptoms were significantly more common among the self-reported false confessors. The assumption of this approach is that variability in false confessions rates is due, at least in part, to individual differences (e.g., a greater desire to please others and protect one's own self-esteem and motivations to avoid conflict) as well as to situational variables (Fulero, 2010a; Kassin & Gudjonsson, 2004).

Third, experimental studies allow for the examination of false confessions in controlled situations where independent variables may be manipulated while controlling extraneous variables. This experimental control allows for stronger internal validity than the other two methods but reduces the ecological validity. For example, experimenters have examined the number of false confessions in artificial situations where volunteer participants are accused of crashing a computer when they did not (e.g., Kassin & Kiechel, 1996) or of cheating (Russanov, Meissner, Narchet, & Kassin, 2005). Many independent variables can be manipulated to examine their potential effects on false confessions, e.g., use of false-evidence ploys (FEPs). Although a thorough understanding of false confessions will ultimately require integration of all three approaches, the current study focuses only upon the experimental studies.

As the studies increase in number, new problems arise. The proliferation of studies and data can overwhelm policymakers and judges and can result in the well-known phenomenon of battling experts, which can in turn lead to confusion among policymakers, scholars, judges, and, in particular, jurors (see Devenport & Cutler, 2004; Lovett & Kovera, 2008). One expert could cite a study with no false confessions (e.g., Hill, Memon, & McGeorge, 2008), and another could cite a study in which all participants falsely confess (e.g., Nash & Wade, 2009).

Several narrative reviews exist for this area of scholarship. For example, Kassin et al. (2010) reviewed theories and research findings from social psychology as well as several experimentally assessed situational factors that can lead to false confession, and they emphasized the isolation and stress that suspects face in police custody, FEPs, plausibility of the forbidden act (discussed subsequently), and minimization. Similarly, Drizin and Leo (2004) reviewed existing studies as well as the social psychological pressures inherent in custodial isolation, presentation of minimizing themes to justify the crime, minimization and maximization, implicit threats and promises, and FEPs. Kassin and Gudjonsson (2004) review similar factors, including plausibility (discussed as “vulnerability,” p. 54), minimization, and the process of confrontation (including interruptions of the suspect’s denials and presentation of fabricated evidence). These reviews and others (Leo, 2008; Woody et al., 2011) reach similar conclusions, particularly regarding the increased risks of false confessions after FEPs, minimization, and confrontation about plausible rather than implausible acts, but they connect these claims to individual studies or cases rather than a systematic meta-analysis of the experimental literature. Due to the potential for traditional narrative reviews to be biased and inaccurate, there has been a general call for the use of systematic reviews and meta-analyses (Blumenthal, 2007; Hunter & Schmidt, 1996).

Meta-analyses can serve two functions beyond narrative reviews. One function is to summarize the results of a body of quantitative research, and a second is to identify weaknesses and gaps in the research. Both of these goals can inform scholars, expert witnesses, and courts about the current state of the field. These goals are particularly relevant in this area for several reasons. First, courts have reached different decisions about the relevance of expert testimony and about the testimony itself in cases involving disputed confessions (see Citron & Johnson, 2006; Fulero, 2010b; Watson, Weiss, & Pouncey, 2010). Second, through neglect, unconscious bias, or deliberation, advocates and experts can choose the studies that support their position and neglect contradicting evidence, and this could affect triers of fact. Third, scholars have demonstrated the biasing effects of expert testimony about interrogation and confession on jurors’ perceptions of the defendant as well as outcomes of simulated trials (e.g., Blandon-Gitlin, Sperry, & Leo, 2011; Gomes, Stenstrom, & Calvillo, 2014; Leo & Liu, 2009; Woody & Forrest, 2009). It is the purpose of
the current study to conduct a meta-analysis of the body of experimental research on false confessions. Specifically, we examine the false confession rates in experimental studies and moderator variables that affected these rates.

Although there are many current variations of experimental methods in the study of the etiology of false confessions, all have the following components. First, participants agree to participate in a psychological experiment in cognition, which is actually a deception. Second, experimenters tell the participants rules for participating in the experiment. Third, the participants are accused of violating the rules. In some studies (i.e., those using the typing or Alt-key method and the individual cheating method, discussed later), all participants are assumed to be innocent and to have not violated the rules. In some studies (e.g., Horselenberg et al., 2006; Horselenberg, Merckelbach, & Josephs, 2003), scholars verified that no participant hit the forbidden key, although this remains unverified in some other Alt-key studies, forming an important limitation of the body of studies using this method. In other experimental methods (i.e., those that use the social cheating method, discussed later), experimenters allow each participant the opportunity to choose whether or not to violate the rules and then accuse all suspects of the rule violation; in this way scholars can compare responses and perceptions of factually guilty and factually innocent suspects (see Houston, Meissner, & Evans, 2014; Meissner et al., 2014). Fourth, the participants are asked to confess to violating the rules. Whether a suspect confesses acts as the dependent variable in the research, and, as a function of the method, these confessions could all be false (i.e., for the Alt-key and the individual cheating method in which all participants are presumed to be factually innocent) or could be false or true, as in the social cheating paradigm.

Many variations on this basic design appear in the literature in order to examine variables that may influence the false confession rate. Variations exist in the type of task, the manner in which participants are asked to confess, and the type of rules given to the participants. Additional components have been added to many studies, such as FEPs (i.e., false claims to have evidence of guilt), the plausibility of the rule violations, and the severity of the consequences for making a false confession. Some of these additions are unique to a study, and others have been used in multiple studies. In the current meta-analysis, we evaluate the following moderators that have been examined frequently enough to be included in a meta-analysis: methods, plausibility, and FEPs.

### 1.1 Experimental methods

The first moderator variable is the experimental method used to study false confessions. Conclusions drawn by experts from such a diverse body of work remain vulnerable to challenges in court. What are the consequences of these methodological differences? To study interrogation under controlled experimental conditions, researchers have sought ethical and realistic methods to simulate interrogations. The current methods include three distinct categories. The first is a negligence method (the Alt-key method) in which the experimenter accuses participants of inadvertently violating a rule. A second method (the social cheating method) involves a confederate who asks the participant to cheat. Regardless of whether the participant actually cheats, the experimenter later accuses the participant of cheating. The third method (the individual cheating method) accuses an individual of cheating without the collaboration of another individual. In each method, the dependent variable was whether or not the participant falsely confessed to the accused act. We briefly review each paradigm.

Kassin and Kiechel (1996) conducted the classic Alt-key study, the first method in this line of inquiry, in which a confederate was to read aloud a list of letters and the participant was to type these letters on the keyboard after being told not to press the Alt-key during the study or the computer would crash. The computer crashed. The researcher then falsely accused the participant of pressing the Alt-key and attempted to get the participant to sign a confession, which would have the additional consequence of “a phone call from the principal investigator” (p. 126). All participants were believed to be innocent. Scholars have since replicated and extended this paradigm to assess the effects of

---

1Perillo and Kassin (2011) excluded a single participant who hit the Alt-key (p. 329).
suspicion that is still substantially less severe than criminal penalties (e.g., payment of
(i.e., a phone call from the presumably upset principal investigator; Kassin & Kiechel, 1996) to more costly conse-
those which may be mild even given the expectations for ethical treatment of participants in psychology research
punishments, the consequences for confessing to pressing the Alt
decision is also expected to be intentional and readily memorable, unlike the Alt
cheating methods, a suspect must decide whether to break the rules, and the decision to follow or break the rules
may confess on the assumption that he or she could have unintentionally hit the Alt
in the Alt
inappropriately helped the confederate) and innocent suspects (i.e., those who appropriately refused to help the con-
federate) during interrogation. The two available meta-analyses of confession studies (Houston et al., 2014; Meissner
et al., 2014) evaluated only six and 12 studies, respectively. For Houston et al. (2014), all included studies employed
only the Russano method and allowed comparison of true confessions and false confessions. More specifically, these
meta-analyses evaluated participants' perceptions of the factors that motivated their true or false confessions in
response to an accusation of cheating. Meissner et al. (2014) required that studies manipulate the interviewing
approach and evaluate both information-gathering and accusatorial methods, and this meta-analysis included some
studies using the Alt-key method as well as studies using the social cheating paradigm. As described subsequently,
we extend meta-analytic techniques to the larger body of experimental scholarship.

A third method, the individual cheating paradigm, was used in only two studies (Horselenberg et al., 2006; Nash
& Wade, 2009). Here individuals worked alone on a task. The experimenter then told each participant that they had
evidence that he or she had cheated. Horselenberg et al. (2006) employed physical evidence to indicate that
participants had looked at the answers to the test, and in the Nash and Wade (2009) study, either the participants
were informed that video evidence existed to show them cheating or they were shown a doctored video that
depicted them cheating. In two of the conditions in the Nash and Wade study it was possible that the participants
could believe they had cheated by accident, and these conditions may be closer to the Alt-key condition than to the
social cheating task.

One of the main differences between the methods is that in both cheating paradigms the participants are
accused of a crime that requires intent, whereas the Alt-key method does not require intent (Houston et al.,
2014). In the Alt-key method, an admission is more similar to admissions of negligence in civil court (see Kassin
& Kiechel, 1996; Perillo & Kassin, 2011), even though actual civil cases involve distinct and complex decision rules
different from those used by participants in these studies (see, e.g., Greene & Bornstein, 2003). Additionally, in
civil disputes, potential coercion may occur not in police interrogation rooms but during depositions or cross-
examination, in which attorneys may seek to inspire an individual to admit that an act of potential negligence
was plausible. Similar questions may arise in criminal courts in jurisdictions that have criminal negligence statutes,
even if there exist many differences between civil and criminal negligence (Garfield, 1998). The Alt-key method
may also share important characteristics with some specific police interrogation tactics. The minimization of intent
in the Alt-key studies is similar to minimization tactics often used by police interrogators who seek to inspire a
suspect to confess to committing the crime accidentally rather than intentionally – a theme that Jayne and
Buckley (1999) call "The Accident Scenario" (p. 470; see also Kassin et al., 2007; Leo, 1996). In contrast, for both
cheating methods, a suspect must decide whether to break the rules, and the decision to follow or break the rules
provides a stronger analog to an actual suspect's decisions regarding whether to commit a criminal act. This
decision is also expected to be intentional and readily memorable, unlike the Alt-key method in which a participant
may confess on the assumption that he or she could have unintentionally hit the Alt-key and either not realized or
forgotten that he or she did so (Houston et al., 2014). These differences make the cheating methods stronger
analogos to the decisions suspects face in actual police interrogations.

Another limitation to the Alt-key method as an analog for police interrogation is that, in comparison to criminal
punishments, the consequences for confessing to pressing the Alt-key are mild. The consequences range from
those which may be mild even given the expectations for ethical treatment of participants in psychology research
(i.e., a phone call from the presumably upset principal investigator; Kassin & Kiechel, 1996) to more costly conse-
quences that are still substantially less severe than criminal penalties (e.g., payment of €250; Horselenberg et al.,
2006). There exist other differences between these methods with less clear legal implications. One such difference is that in the Alt-key method the participants actually see the effect of rule violation, i.e., the computer crashes, while in the social cheating paradigm they are only told that a violation occurred. Another difference is that in the social cheating paradigm a conscious collaboration is required, but the Alt-key method and the individual cheating method do not require collaboration among individuals.

1.2 Plausibility

Scholars using the Alt-key paradigm have suggested that the likelihood of committing a false confession is related to the plausibility of the forbidden act (Kassin & Kiechel, 1996). Plausibility has been manipulated in two ways, both of which affect the perceived likelihood of erroneously striking the Alt-key. First, researchers manipulated the speed at which the stimuli appeared on the screen, with the hypothesis that individuals would be more likely to make a false confession because the higher rate of presentation made errors more plausible (Kassin & Kiechel, 1996). The second method varies the location of the forbidden key. If the forbidden key is close to the response key (e.g., the Alt-key next to the spacebar), then hitting it would be more plausible than if it was far away from the response key (e.g., the Esc-key far from the spacebar). Both methods seek to influence participants’ “subjective uncertainty concerning their own innocence” (Kassin & Kiechel, 1996, p. 126).

1.3 False-evidence ploys

Another variable manipulated in a number of studies was the use of FEPs, the presentation of fabricated evidence indicating that the participant committed the crime. Within the Alt-Key paradigm this typically consisted of a witness falsely claiming that he or she saw the participant press the forbidden key (see, e.g., Kassin & Kiechel, 1996). In the cheating paradigm, the FEPs have included explicit or implicit claims to have video evidence of cheating (e.g., Perillo & Kassin, 2011). Across all methods, the hypothesis has been that individuals are more likely to confess falsely when confronted with fabricated evidence of their guilt (see Kassin et al., 2010; Leo, 2008; Woody et al., 2011; for reviews of effects of FEPs).

1.4 Overview of the current study

To understand better the research literature on false confessions, we assessed previously conducted false confession research with meta-analysis. Application of meta-analysis of false confession research can enhance our understanding of false confession rates across methodologies. Furthermore, we can examine similarities and differences between and within studies to evaluate the effects that particular moderator variables have on false confession rates.

As stated previously, results from this and other meta-analyses can inform and support experts’ testimony in court. Although there exists broad agreement among many scholars about false confessions and about some of the moderators listed here (e.g., FEPs; see Kassin, 2008; Kassin et al., 2010; Kassin & Gudjonsson, 2004; Leo, 2008; Woody et al., 2011), larger agreement through the criminal justice system remains elusive. As Cutler, Findley, and Loney (2014) stated, “The courts’ response to expert testimony on false confessions ... has not been uniformly welcoming” (p. 590). As reported by Fulero (2010a, 2010b), courts may reject psychological testimony due to questions of scientific validity or reliability as well as for other reasons related to Frye, Daubert, or other criteria for admission of experts (see also Cutler et al., 2014; Watson et al., 2010). Inferences about reliability rest on claims testable via meta-analysis.

Additionally, John E. Reid and Associates, Inc. (2010) list more than 60 cases from 2002–2010 in which courts excluded, limited, or rejected expert testimony [they defined “rejected” to include “cases in which the expert offers some testimony but their [sic] argument was rejected by the judge or jury” (p. 6)]. They then take a stronger stance as they summarize a series of court decisions by stating that, “For the past several years the courts have viewed with skepticism the testimony of ‘false confession experts,’ [sic] repeatedly suggesting that there is no actual
science to support their views but rather, anecdotal evidence” (John E. Reid and Associates, 2015a, ¶5). Although these authors are practitioners rather than scholars and although they have a long-term financial interest in the success of the Reid Technique in general and FEPs in particular (see Inbau, Reid, Buckley, & Jayne, 20112), they have trained thousands of police investigators, and they present extensive materials for prosecutors and others to challenge the testimony of confessions in court. In this meta-analysis, we seek to bring more clarity to the ongoing disparities between reliable findings commonly accepted by scholars and rejection of those findings by courts, trainers of interrogators, and others.

Broadly speaking there are two forms of meta-analysis. The first, and most common in psychology, synthesizes relationships among variables within studies. For example, as discussed previously, one meta-analysis examined differences in psychological process in true and false confessions, using six studies that compared both types of confessions (Houston et al., 2014). Such meta-analyses usually focus on correlations or differences among groups. Another meta-analysis (Meissner et al., 2014) examined how different interrogation techniques influenced the false confession and included only studies that compared interrogation techniques.

The second form of meta-analysis focuses on a single variable within a study and examines the estimate of the variable across studies and how it changes with selected moderator variables (Lipsey & Wilson, 2000). This form of meta-analysis includes such foci as test reliability and prevalence studies. Prevalence-based meta-analysis focuses on the proportion of an event within a study (Barendregt, Doi, Lee, Norman, & Vos, 2013). In prevalence-based meta-analysis, relations among variables are examined across studies. Simply stated, one study represents one level of a moderator and another study represents another level of the moderator variables. Such between-studies analyses, because they lack random assignment, have more uncertainty than when the moderator variables are examined within studies that employ random assignment. Nevertheless, such analyses are suggestive of meaningful relations that may be investigated in future research (Wood & Eagly, 2009). One can examine relations among variables that have never been investigated within any single study. Although such an examination may lack the controls of within-study research, they establish the feasibility of hypotheses that can be investigated in future research (Wood & Eagly, 2009).

We employed a prevalence meta-analysis using the false confession rate as the effect size. The false confession rate is the number of participants who make a false confession out of the total number of participants. We compare the false confessions rate across studies and within studies when possible. In order to combine effect sizes across and within studies, we treat independent groups within a study as separate observations and do not employ the statistical analysis conducted within the studies.

2 | METHOD

2.1 | Identification, retrieval, and selection of studies

Using relevant individual search terms (i.e., “false admission,” “false confession,” “false evidence,” “laboratory confession,” “interrogation,” and “Alt-key”), we searched the following databases for relevant studies: Criminal Justice Abstracts, ERIC, CSA Linguistics and Language Behavior Abstracts, National Criminal Justice Reference Service Abstracts, PAIS International, PsycARTICLES, Social Services Abstracts, Sociological Abstracts, PsycINFO, Access Science, Criminal Justice Abstracts, Criminal Justice Periodicals, Dissertations & Theses, Dissertations & Theses: A&I, Dissertations & Theses: Full Text, JSTOR, GOOGLE scholar, Social Sciences Full Text, Science Citation Index and Science Direct. From this search, we selected experimental studies conducted in laboratory settings in which researchers attempted to elicit confessions from adult participants after falsely accusing the participants of

2For a historical point, the Inbau and Reid (1967) manual stated, “if deceit is impermissible a ban will have to be placed on all interrogations of criminal suspects. And without some elements of ‘trickery,’ such as leading the suspect to believe that the police have some tangible or specific evidence of his guilt, many interrogations will be totally ineffective” (pp. 196–197).
committing a forbidden act. We found additional references by searching the Science Citation Index for relevant articles that cited the selected articles and by examining the references cited in each of the selected studies. We found 19 publications that fitted our criteria for inclusion.

2.2 Coding system and moderator variables

We developed a coding system to rate each data point within the studies. For the data points in this meta-analysis, we used the individual experimental groups within the studies. The data points consisted of non-overlapping participants who experienced different experimental manipulations; therefore, no participants were in more than one data point. We focused on data points, rather than averaging within a study, to examine the effects of moderator variables.

For each data point, we coded the number of confessions and the total number of participants. Based upon this information, we calculated the prevalence rate of false confessions. We also coded the following potential moderator variables: (1) experiment method (Alt-key, social cheating, or individual cheating); (2) plausibility-typing speed (fast or slow); (3) plausibility-location (high or low); and (4) FEP (present or absent). When coding typing speed, we coded a presentation rate of > 60 characters/minute as fast, and we coded a presentation rate of < 60 characters/minute as slow. The median for fast typing was 67 characters/minute (range 64–67; all but one study had a fast typing rate of 67 characters/minute), and the median for slow typing was 43 characters/minute (range 43–55; all but two studies had a slow typing rate of 43 characters/minute). We coded plausibility-location for typing tasks only, and we coded plausibility-location related to the placement of the forbidden key. We coded forbidden keys within two rows of keys from the spacebar as having high plausibility of error (e.g., Alt-key), and we coded forbidden keys more than two rows away from the spacebar as having low plausibility of error (e.g., Esc-key).

Several studies failed to report the actual number of participants in each data point but reported the percentage of confessions in each. We contacted authors of these articles, but some did not respond to repeated requests. Therefore, in several studies we estimated the number of participants included in each condition by assuming that the groups were all the same size.

2.3 Analysis

We employed Comprehensive Meta Analysis 2.0 (Borenstein, Hedges, Higgins, & Rothstein, 2005) and Meta Win 2.0 (Rosenberg, Adams, & Gurevitch, 2000) as the statistical platforms for completing the statistical analysis. For the summation of the prevalence findings, we computed prevalence point estimates and 95% confidence intervals (CIs).

The prevalence effect size was the event rate (i.e., number of individuals in a group making a false confession divided by the total number of persons in the group). After a continuity correction, we converted each event rate to a log event rate for the analysis, and we converted it back to event rate when reporting the results. We examined the results with other data transformations (see Barendregt et al., 2013), and the results with other transformations were quite similar for this dataset.

We used a random effects model to calculate the overall mean effect sizes and the 95% CIs (Borenstein, Hedges, Higgins, & Rothstein, 2009). Random effects models allow for the possibilities that there exist both errors stemming from sampling error and random differences between studies associated with variations in procedures, settings, and similar factors. The random effects model was more plausible because data points with different experimental manipulations served as our unit of analysis.

We conducted moderator analyses with a mixed effects model. In a mixed effects analysis, a random effects model is used to combine studies within each group to be compared. A fixed effects model is then used across the groupings in the moderator variables. Here we used a fixed effects model due to the small number of levels within the moderator variables. We based all moderator variables in this analysis upon dichotomous groupings of the data points.
RESULTS

3.1 Database demographics

The resulting database for this meta-analysis consisted of 19 publications or theses with 24 studies; some publications included more than one study, and 52 data points contained a total of 1,638 participants (Tables 1 and 2). The participants in these studies were not representative of the population of the United States or of defendants.

TABLE 1 Typing task studies by first author and year, with sample size, false confession rate (FCR), plausibility location keys, typing speed, and presence (yes or no) of false-evidence ploy (FEP)

<table>
<thead>
<tr>
<th>Study</th>
<th>Year</th>
<th>Sample size</th>
<th>FCR rate</th>
<th>95% CI</th>
<th>Key(s)</th>
<th>FEP</th>
<th>Typing speed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blair a</td>
<td>2007</td>
<td>98</td>
<td>0.27</td>
<td>0.19–0.36</td>
<td>Alt-Ctrl-Del</td>
<td>Yes</td>
<td>Not reported</td>
</tr>
<tr>
<td>Blair b</td>
<td>2007</td>
<td>98</td>
<td>0.29</td>
<td>0.21–0.38</td>
<td>Alt-Ctrl-Del</td>
<td>No</td>
<td>Not reported</td>
</tr>
<tr>
<td>Forrest a</td>
<td>2002</td>
<td>27</td>
<td>0.52</td>
<td>0.33–0.70</td>
<td>Alt</td>
<td>No</td>
<td>67/minute</td>
</tr>
<tr>
<td>Forrest b</td>
<td>2002</td>
<td>29</td>
<td>0.69</td>
<td>0.50–0.83</td>
<td>Alt</td>
<td>No</td>
<td>67/minute</td>
</tr>
<tr>
<td>Forrest</td>
<td>2006</td>
<td>98</td>
<td>0.82</td>
<td>0.72–0.88</td>
<td>Alt</td>
<td>No</td>
<td>67/minute</td>
</tr>
<tr>
<td>Hickcox</td>
<td>2009</td>
<td>56</td>
<td>0.75</td>
<td>0.62–0.85</td>
<td>Alt</td>
<td>No</td>
<td>Not reported</td>
</tr>
<tr>
<td>Horselenberg 1a</td>
<td>2006</td>
<td>26</td>
<td>0.58</td>
<td>0.39–0.75</td>
<td>F12</td>
<td>Yes</td>
<td>Not reported</td>
</tr>
<tr>
<td>Horselenberg 1b</td>
<td>2006</td>
<td>30</td>
<td>0.77</td>
<td>0.59–0.88</td>
<td>Windows</td>
<td>Yes</td>
<td>Not reported</td>
</tr>
<tr>
<td>Horselenberg 2</td>
<td>2006</td>
<td>9</td>
<td>0.11</td>
<td>0.02–0.50</td>
<td>Windows</td>
<td>Yes</td>
<td>Not reported</td>
</tr>
<tr>
<td>Horselenberg 3</td>
<td>2003</td>
<td>34</td>
<td>0.79</td>
<td>0.62–0.92</td>
<td>Shift</td>
<td>Yes</td>
<td>Not reported</td>
</tr>
<tr>
<td>Kassin a</td>
<td>1996</td>
<td>17</td>
<td>0.97</td>
<td>0.68–0.99</td>
<td>Alt</td>
<td>Yes</td>
<td>67/minute</td>
</tr>
<tr>
<td>Kassin b</td>
<td>1996</td>
<td>18</td>
<td>0.89</td>
<td>0.65–0.77</td>
<td>Alt</td>
<td>Yes</td>
<td>43/minute</td>
</tr>
<tr>
<td>Kassin c</td>
<td>1996</td>
<td>17</td>
<td>0.65</td>
<td>0.40–0.83</td>
<td>Alt</td>
<td>No</td>
<td>67/minute</td>
</tr>
<tr>
<td>Kassin d</td>
<td>1996</td>
<td>23</td>
<td>0.26</td>
<td>0.12–0.47</td>
<td>Alt</td>
<td>No</td>
<td>43/minute</td>
</tr>
<tr>
<td>Klaver a</td>
<td>2008</td>
<td>71</td>
<td>0.70</td>
<td>0.59–0.80</td>
<td>Alt</td>
<td>No</td>
<td>67/minute</td>
</tr>
<tr>
<td>Klaver b</td>
<td>2008</td>
<td>62</td>
<td>0.47</td>
<td>0.35–0.59</td>
<td>Alt</td>
<td>No</td>
<td>67/minute</td>
</tr>
<tr>
<td>Klaver c</td>
<td>2008</td>
<td>30</td>
<td>0.23</td>
<td>0.12–0.42</td>
<td>Esc</td>
<td>No</td>
<td>67/minute</td>
</tr>
<tr>
<td>Klaver d</td>
<td>2008</td>
<td>40</td>
<td>0.05</td>
<td>0.01–0.18</td>
<td>Esc</td>
<td>No</td>
<td>67/minute</td>
</tr>
<tr>
<td>Newring</td>
<td>2008</td>
<td>26</td>
<td>0.19</td>
<td>0.08–0.39</td>
<td>Alt</td>
<td>No</td>
<td>64/minute</td>
</tr>
<tr>
<td>Perillo 1a</td>
<td>2010</td>
<td>14.2</td>
<td>0.27</td>
<td>0.10–0.54</td>
<td>Alt</td>
<td>No</td>
<td>43/minute</td>
</tr>
<tr>
<td>Perillo 1b</td>
<td>2010</td>
<td>14.2</td>
<td>0.36</td>
<td>0.16–0.62</td>
<td>Alt</td>
<td>No</td>
<td>43/minute</td>
</tr>
<tr>
<td>Perillo 1c</td>
<td>2010</td>
<td>14.2</td>
<td>0.29</td>
<td>0.16–0.62</td>
<td>Alt</td>
<td>No</td>
<td>43/minute</td>
</tr>
<tr>
<td>Perillo 1d</td>
<td>2010</td>
<td>14.2</td>
<td>0.87</td>
<td>0.59–0.97</td>
<td>Alt</td>
<td>Yes</td>
<td>43/minute</td>
</tr>
<tr>
<td>Perillo 1e</td>
<td>2010</td>
<td>14.2</td>
<td>0.77</td>
<td>0.50–0.92</td>
<td>Alt</td>
<td>Yes</td>
<td>43/minute</td>
</tr>
<tr>
<td>Perillo 2a</td>
<td>2010</td>
<td>19</td>
<td>0.74</td>
<td>0.51–0.89</td>
<td>Alt</td>
<td>Yes</td>
<td>43/minute</td>
</tr>
<tr>
<td>Perillo 2b</td>
<td>2010</td>
<td>19</td>
<td>0.47</td>
<td>0.27–0.69</td>
<td>Alt</td>
<td>No</td>
<td>43/minute</td>
</tr>
<tr>
<td>Redlich a</td>
<td>2003</td>
<td>16</td>
<td>0.69</td>
<td>0.43–0.86</td>
<td>Alt</td>
<td>No</td>
<td>55/minute</td>
</tr>
<tr>
<td>Redlich b</td>
<td>2003</td>
<td>16</td>
<td>0.50</td>
<td>0.27–0.73</td>
<td>Alt</td>
<td>Yes</td>
<td>55/minute</td>
</tr>
<tr>
<td>Swanner a</td>
<td>2010</td>
<td>25</td>
<td>0.65</td>
<td>0.45–0.81</td>
<td>Tab</td>
<td>Yes</td>
<td>67/minute</td>
</tr>
<tr>
<td>Swanner b</td>
<td>2010</td>
<td>25</td>
<td>0.62</td>
<td>0.42–0.78</td>
<td>Tab</td>
<td>Yes</td>
<td>67/minute</td>
</tr>
<tr>
<td>Swanner c</td>
<td>2010</td>
<td>25</td>
<td>0.52</td>
<td>0.33–0.70</td>
<td>Tab</td>
<td>No</td>
<td>67/minute</td>
</tr>
<tr>
<td>Swanner d</td>
<td>2010</td>
<td>25</td>
<td>0.26</td>
<td>0.12–0.46</td>
<td>Tab</td>
<td>No</td>
<td>67/minute</td>
</tr>
</tbody>
</table>

Note: FEP is either present (yes) or absent (no). The letters following each first author’s name indicate experimental subgroups of distinct participants. Sample sizes that are not whole numbers reflect estimates based on percentages reported in publications. CI, confidence interval.
in criminal cases, as discussed later. All studies evaluated undergraduate students as participants; when specified, all students were in the social sciences. Based upon those studies that reported age, the mean age was 19.60 years (range 18–21.2). All but one study reported the number of males and females; overall, males comprised slightly less than one-third of the sample (32.56%). Other demographic characteristics (e.g., ethnicity) were reported in only a few studies. Comparisons of gender and ethnicity remained rare.

### 3.2 Mean effect sizes

The point estimate event rate across groups was 0.47 (95% CI: 0.40–0.55). Across all studies, 47% of the participants falsely confessed. Although our *a priori* decision was to use the random effects model, we examined the fixed effects model as well. The overall rate and CI were similar to the random effects model (fixed model point estimate = 0.47, 95% CI: 0.44–0.50), but the fit of the fixed effects model was poor ($Q = 321.01$, $df = 51$, $p < 0.001$), and $I^2$ was 84.11, suggesting a very large degree of heterogeneity that cannot be explained only by sampling error (Higgins, Thompson, Deeks, & Altman, 2003).

In any meta-analysis, there is always the possibility that studies exist that were not included due to unavailability and/or being unpublished (i.e., the file-drawer effect). We employed Orwin’s (1983) method to determine the number of studies that would make the prevalence trivial. We found that 89 studies with a 1 per 100 prevalence would have to exist to bring the overall prevalence down to 5 per 100.

Publication and selection biases in meta-analysis are more likely to be found in studies with small sample sizes than in those with large sample sizes (Begg, 1994). One way of examining whether that occurred in the current

---

### TABLE 2 Cheating task studies by first author and year, with sample size, confession rate, and cheating type (social or individual)

<table>
<thead>
<tr>
<th>Study</th>
<th>Year</th>
<th>Sample size</th>
<th>Confession rate</th>
<th>95% CI</th>
<th>Cheating type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Guyll</td>
<td>2013</td>
<td>74</td>
<td>0.43</td>
<td>0.33–0.55</td>
<td>Social</td>
</tr>
<tr>
<td>Hill a</td>
<td>2008</td>
<td>14</td>
<td>0.03</td>
<td>0.00–0.37</td>
<td>Social</td>
</tr>
<tr>
<td>Hill b</td>
<td>2008</td>
<td>13</td>
<td>0.04</td>
<td>0.00–0.38</td>
<td>Social</td>
</tr>
<tr>
<td>Horgan a</td>
<td>2009</td>
<td>33</td>
<td>0.42</td>
<td>0.27–0.60</td>
<td>Social</td>
</tr>
<tr>
<td>Horgan b</td>
<td>2009</td>
<td>33</td>
<td>0.21</td>
<td>0.11–0.38</td>
<td>Social</td>
</tr>
<tr>
<td>Marschall</td>
<td>2013</td>
<td>73</td>
<td>0.29</td>
<td>0.19–0.40</td>
<td>Social</td>
</tr>
<tr>
<td>Narchet a</td>
<td>2005</td>
<td>35</td>
<td>0.20</td>
<td>0.10–0.36</td>
<td>Social</td>
</tr>
<tr>
<td>Narchet b</td>
<td>2005</td>
<td>35</td>
<td>0.40</td>
<td>0.25–0.57</td>
<td>Social</td>
</tr>
<tr>
<td>Narchet c</td>
<td>2005</td>
<td>35</td>
<td>0.29</td>
<td>0.16–0.45</td>
<td>Social</td>
</tr>
<tr>
<td>Perillo 3a</td>
<td>2010</td>
<td>15</td>
<td>0.03</td>
<td>0.00–0.35</td>
<td>Social</td>
</tr>
<tr>
<td>Perillo 3b</td>
<td>2010</td>
<td>15</td>
<td>0.50</td>
<td>0.27–0.73</td>
<td>Social</td>
</tr>
<tr>
<td>Russano a</td>
<td>2005</td>
<td>37</td>
<td>0.05</td>
<td>0.01–0.20</td>
<td>Social</td>
</tr>
<tr>
<td>Russano b</td>
<td>2005</td>
<td>37</td>
<td>0.14</td>
<td>0.06–0.29</td>
<td>Social</td>
</tr>
<tr>
<td>Russano c</td>
<td>2005</td>
<td>37</td>
<td>0.19</td>
<td>0.09–0.35</td>
<td>Social</td>
</tr>
<tr>
<td>Russano d</td>
<td>2005</td>
<td>37</td>
<td>0.43</td>
<td>0.28–0.60</td>
<td>Social</td>
</tr>
<tr>
<td>Horselenberg 3</td>
<td>2006</td>
<td>16</td>
<td>0.06</td>
<td>0.01–0.41</td>
<td>Individual</td>
</tr>
<tr>
<td>Nash 1a</td>
<td>2009</td>
<td>15</td>
<td>0.73</td>
<td>0.47–0.90</td>
<td>Individual</td>
</tr>
<tr>
<td>Nash 1b</td>
<td>2009</td>
<td>15</td>
<td>0.97</td>
<td>0.65–0.99</td>
<td>Individual</td>
</tr>
<tr>
<td>Nash 2a</td>
<td>2009</td>
<td>15</td>
<td>0.80</td>
<td>0.53–0.93</td>
<td>Individual</td>
</tr>
<tr>
<td>Nash 2b</td>
<td>2009</td>
<td>15</td>
<td>0.93</td>
<td>0.65–0.99</td>
<td>Individual</td>
</tr>
</tbody>
</table>

*Note:* The letters following each first author’s name indicate experimental subgroups of distinct participants. CI, confidence interval.
meta-analysis is the Begg and Mazumdar rank correlation method (Begg & Mazumdar, 1994), which examines the relationship between the ranks assigned for the effect size and the sample sizes. Significant correlations suggest a publication bias. In the current study, we found a very low rank correlation (τ = 0.05, p = 0.54). Similar results were found with Egger's test (t = 0.06, p = 0.95; df = 50).

To estimate what the weighted average effect size might be without publication bias, we used the trim-and-fill method (Duval & Tweedie, 2000). The funnel plot showed no obvious signs of asymmetry, and no adjustment was necessary.

3.3 | Sensitivity analyses

We examined the robustness of the overall findings against the effects of an outlier by sequentially eliminating one study at a time and reanalyzing the remaining dataset. A new analysis was conducted after each study was removed. Even when the study with the largest impact was dropped, the resulting point estimate changed by only 0.02. The impact of any single study remained trivial.

We decided to evaluate the effects of data points with no confessions or all confessions. To check how our decision process affected the data, we examined the prevalence after dropping these five studies. The results were very similar to the total sample; the point estimate was 0.48 (95% CI: 0.40–0.56) for the random effects model.

3.4 | Moderator analyses

The only moderator analysis to employ the entire dataset was the comparison of the experimental task (Alt-key, social cheating, or individual cheating). The other moderator analyses examined subsets of studies within the three task types.

3.4.1 | Moderator variables

There was a significant difference across the three methods (Q = 21.54, df = 2, p < 0.001). The event rate (i.e., prevalence of false confessions) of the 32 Alt-key data points (see Table 1) was 0.55 (95% CI: 0.46–0.64). The 15 social cheating data points (see Table 2) had a mean false confession rate of 0.26 (95% CI: 0.18–0.3), and the five individual cheating data points (see Table 2) had a mean false confession rate of 0.75 (95% CI: 0.48–0.90). The event rate for the social cheating method was significantly lower than for the other two methods, which did not differ from each other (social cheating vs. Alt-key, Q = 16.70, df = 1, p < 0.001; social cheating vs. individual cheating, Q = 14.91, df = 1, p < 0.001; Alt-key vs. individual cheating, Q = 2.05, df = 1, p = 0.15). Due to the differences across methods, we examined the remaining moderators within the Alt-key task.

3.4.2 | Within Alt-key task moderators: Plausibility-typing speed

Kassin and Kiechel (1996) hypothesized that asking participants to type faster would lead to greater rates of false confessions due to their perceived increased likelihood of accidentally pressing the forbidden key. In spite of the initial findings of significant differences (Kassin & Kiechel, 1996), across studies speed at which the typing task was performed did not serve as a significant predictor of false confession rates (Q = 0.95, df = 1, p = 0.33). The typing of 11 data points using a slower typing rate had a mean confession rate of 0.62 (95% CI: 0.45–0.76), and the 14 data points with a fast-paced typing rate had a mean confession rate of 0.51 (95% CI: 0.38–0.64).

3.4.3 | Within Alt-key task moderator: Plausibility-location

In several studies, researchers hypothesized that placing the forbidden key far from the spacebar made false confessions less likely because of implausibility of making such errors. Manipulation of plausibility with the placement of the forbidden key in the typing paradigm studies had a significant effect on the false confession rate (Q = 10.01, df = 1, p = 0.01). For the 23 data points with forbidden keys close to the spacebar (e.g., Alt-key, Shift-key, Windows-key or
Control-key) the mean false confession rate was 0.64 (95% CI: 0.55–0.72), and the false confession rate for the 9 data points with the forbidden key was placed far from the spacebar (e.g., Esc-key) was 0.37 (95% CI: 0.25–0.51). Interestingly, the only study to evaluate plausibility-location directly (Horselenberg et al., 2006) found no significant difference.

### 3.4.4 Within Alt-key task moderators: FEPs

Several studies hypothesized that the use of FEPs would increase the false confession rate. Across all typing data points we found that FEPs elicited significantly higher rates of false confession ($Q = 4.70, \text{df} = 1, p = 0.030$). The mean false confession rate for the 14 data points with an FEP was 0.68 (95% CI: 0.54–0.79), and for the 18 data points without an FEP the mean false confession rate was 0.47 (95% CI: 0.35–0.59).

A secondary meta-analysis of the six Alt-key studies that directly compared FEPs with no FEPs found that FEPs produced more confessions (log odds $=1.034$, $p = 0.044$). Across these studies, the false confession rate with FEPs was 57%, and that without FEPs was 37%.

In the individual cheating studies, all conditions used FEPs, while the social cheating studies did not. The single social cheating study that compared FEP conditions (i.e., explicit FEP and bluff, a form of implicit FEP) with a no-FEP condition and found significantly more confessions after FEPs (Perillo & Kassin, 2011).

### 4 DISCUSSION

Across studies, approximately 47% of innocent participants falsely confessed. This effect does not differentiate between methods; rather it serves as a limited representation of false confession rates among experimental studies. Additional cautions are necessary. As discussed previously, experimental studies bring benefits of extensive experimental control and ethical treatment of participants, and these factors reduce realism. Additionally, unlike police investigators, researchers have access to the ground truth, and many of the studies in this meta-analysis evaluated only participants presumed to be innocent. Therefore, this single prevalence number reflects the structures of interrogation research rather than an estimated false confession rate among actual suspects, and we do not attempt to connect this outcome to any estimates of the actual rate of false confessions in police interrogations, which, as discussed previously, appear substantially lower as assessed using a variety of methods. In the materials that follow, we evaluate moderator variables and the ways in which these meta-analytic outcomes can guide research and practice.

#### 4.1 Task

The task served as a major differentiating moderator variable between studies. Studies that utilized the Alt-key task resulted in higher rates of false confessions than did the studies that employed social cheating tasks. A fundamental distinction between the Alt-key studies and social cheating methods is the nature of the act to which participants confess. As discussed, a suspect’s admission that he or she accidentally hit the Alt-key is more similar to an admission of negligence than to a confession of an intentional criminal act (Houston et al., 2014; Perillo & Kassin, 2011). These findings suggest that cheating and typing studies comprise distinctive paradigms that reflect the differences between a criminal action and a negligence tort. Future research should seek to examine moderator variables between and within these approaches.

For an additional analysis of the consequences of the task, we compared social and individual cheating methods. The data points from the Russano et al. (2005) social cheating method (e.g., Horgan, 2009; Narchet, 2005; Perillo & Kassin, 2011) resulted in a smaller probability of eliciting a false confession than did studies that used individual cheating tasks (e.g., Hill et al., 2008; Horselenberg et al., 2006; Nash & Wade, 2009). This finding may relate to the power of the FEPs used in all individual cheating studies. For example, across all conditions of the Nash and Wade
(2009) study, all participants confessed when confronted with false claims of video evidence or shown fabricated video evidence. These outcomes demonstrate that differences in interrogation tactics under study and differences in the methods to assess these tactics can lead to important differences in confession rates.

4.2 | Plausibility of error: typing speed and location

Some scholars (e.g., Kassin, 1997; Kassin & Kiechel, 1996) have suggested that the speed at which participants type would influence false confession rates such that participants who type faster would be more likely to believe they made a typing error and, in turn, would be more likely to confess falsely. Similarly, scholars have argued that the location of the forbidden key would affect participants’ certainty about their innocence and that participants accused of making a more plausible error would be more likely to confess falsely. As this meta-analysis demonstrates, typing speed does not affect participants’ likelihood of false confession, but when the forbidden key was near the spacebar, participants’ rates of false confession increased. One explanation for the mixed effects is the small number of conditions that investigated these variables; despite the limitations of the Alt-key method, we recommend a greater degree of replication of research examining typing speed and key location to increase the strength of future meta-analyses.

A second explanation relates to typing speed and forbidden key location as analogs of participants’ uncertainty about their innocence and the applicability of this method to criminal interrogations. As noted previously, the Alt-key task does not simulate confessions of intentional criminal acts (see Houston et al., 2014; Perillo & Kassin, 2011), but we recommend further evaluation of this negligence method for three reasons. First, although these manipulations may not simulate a suspect’s perceptions of the likelihood that he or she committed a major crime, many negligence cases hinge on defendants’ admissions to have failed to take reasonable precautions, and these admissions may function in ways that are similar to the Alt-key paradigm. For example, in a civil or criminal negligence case resulting from an automobile accident, a defendant may be more likely to admit to not using his or her turn signal (i.e., a more plausible act) than to swerving radically across lanes (i.e., a less plausible act). We encourage scholars to extend these research methods into civil questions related to admissions of liability and to resultant damage awards, and we encourage scholars to continue their search for ethically appropriate and ecologically valid manipulations.

A second justification for the continued study of plausibility relates to police interrogation strategies. These findings demonstrate that more plausible false accusations are more likely to induce false confessions, and we encourage scholars to investigate this possibility experimentally as well as with more ecologically valid methods. A third concern relates to the plausibility of FEPs. Alongside concerns about plausible but false accusations, highly plausible forms of FEPs (i.e., fabricated evidence that appears highly plausible), as recommended by trainers of interrogators (Inbau et al., 2011; Jayne & Buckley, 1999), may raise false confession rates. Scholars have not yet experimentally investigated these questions as applied to more or less plausible FEPs.

4.3 | False-evidence ploys

Across studies and methods, FEPs that included presentation of described or actual fabricated evidence increased rates of false confessions.3 These findings parallel concerns about FEPs in actual cases, particularly because FEPs have “been implicated in the vast majority of documented police-induced false confessions” (Kassin et al., 2010, p. 12). Although John E. Reid and Associates (2015b) argue that a limitation of these studies is that many researcher-

---

3A bluff is a form of an implicit FEP (see Woody et al., 2013) in which investigators falsely claim that evidence exists but will be tested in the future (see Perillo & Kassin, 2011). We included examinations of bluffs by Perillo and Kassin (2011) in our analyses of studies that include explicit FEPs (i.e., a ploy in which investigators directly claim that false evidence exists) for three reasons. First, Perillo and Kassin (2011) found that false confession rates from explicit and implicit FEPs did not differ and that these forms of FEPs raised false confession rates in similar ways. Second, Woody et al. (2013) found that jurors do not distinguish between implicit and explicit FEPs. Third, Woody et al. (2013) argue that implicit and explicit FEPs both involve police deception about evidence and that these methods do not function in psychologically distinct ways.
interrogators have not been formally trained in these techniques, some researchers have received training, and, perhaps more importantly, FEPs executed by trained, experienced interrogators may be more rather than less powerful for both guilty and innocent suspects.

These results suggest that investigators and those who train them (e.g., Inbau et al., 2011; Jayne & Buckley, 1999) remain overconfident in claims that innocent suspects remain safe from negative consequences of FEPs. For example, Inbau et al. (2011) ask readers to “consider an innocent rape suspect who is falsely told that DNA evidence implicates his guilt. Would this false statement cause an innocent person to confess? Of course not! Under these conditions [false confession] becomes much more plausible ... – not because fictitious evidence was presented, but because that evidence was used to augment an improper interrogation technique” (pp. 351–352). The authors then refer to the possibility that an FEP alone could cause a false confession as “absurd” (p 352). Inbau et al. (2011) and others (e.g., Jayne & Buckley, 1999) recommend caution with FEPs, not because of the potential for false confession but because an FEP may backfire and inform a suspect that the police do not have actual evidence (e.g., if police falsely claim that a witness saw the suspect flee through the front door, the guilty suspect who knows that he or she fled out the back door now knows the police do not have an actual eyewitness). We suggest far more caution. Rather than being implausible, the stark and consistent findings of increased false confession rates resulting from FEPs provide further support for claims that these forms of deception are coercive for innocent suspects, particularly given findings from Rogers et al. (2010) that a majority of criminal defendants in their sample erroneously believed that police deception about eyewitness evidence is “legally wrong” (p. 310). Additionally, these findings raise important concerns because under controlled experimental conditions it remains highly unlikely that researchers consistently included improper, illegal, or coercive yet unreported interrogation tactics in their methods, as alleged by Inbau et al. (2011). The increase in false confession rates across studies appears to rest directly on the use of FEPs alone.

These findings reinforce some observers’ concerns about the effects of FEPs, particularly because jurors and juries accept confessions and often convict defendants even when confessions result from FEPs (Woody & Forrest, 2009; Woody, Forrest, & Yendra, 2013). Further research, however, should seek to examine how types of FEPs, such as demeanor, testimonial, or scientific FEPs (see Forrest et al., 2012; Leo, 2008; Woody & Forrest, 2009) may differentially affect false confession rates. These meta-analytic findings provide additional justification for courts to revisit the coerciveness of police deception during interrogation, particularly deception about evidence, (see Bandler, 2014a, 2014b; McKinley, 2014; People v. Thomas, 2014) as well as for the calls by some scholars for the United States to follow Great Britain in the elimination of police deception during interrogation (e.g., Kassin et al., 2010; Kassin & Gudjonsson, 2004; Woody et al., 2011).

4.4 | Limitations

Limitations exist both for this meta-analysis and for the experimental study of false confessions. Limits to the present meta-analysis include the lack of access to all raw data and extensive variability of research methods across studies. Researchers often did not include raw data, particularly for prevalence of false confessions within each experimental condition, and, as stated previously, some did not respond to our requests for information; therefore, in some cases we estimated the number of participants in conditions.

The extensive variability in these studies is a strength as well as a limitation of the field because scholars have evaluated a wide range of crimes and interrogation methods that include general techniques (e.g., minimization, which includes reducing the perceived legal seriousness of the crime, reducing the perceived moral seriousness of the crime, implicitly reducing the perceived potential consequences of confession, and similar tactics) as well as more specific interrogation tactics [e.g., FEPs involving physical evidence or fabricated video evidence; see Horselenberg et al.,

4Among other scholars who have been trained, the following have personally verified their training with at least one of the current authors: J. T. Perillo (personal communication, March, 2010) and S. A. Woestehoff (personal communication, September, 2013).

5For a prior use of “absurd,” see the 4th edition of the manual (Inbau, Reid, Buckley, & Jayne, 2001, p. 429).
2006 and Nash & Wade, 2009, respectively]. This variation in research methods and variables, however, allows the field
to reflect the extensive variation in police interrogation tactics (Leo, 2008; Wallace, 2010). This variability, however,
also limits the general conclusions available through meta-analysis. The apparent variability may also be misleading,
however, because many studies in this sample emerged from programmatic research generated by a small number of
laboratories using similar methods with relatively homogenous populations of university students as participants.

Other limitations of this body of research also limit the meta-analytic conclusions. The participants are not
representative, the consequences of confessing are not representative, and the tasks are not representative of actual
criminal interrogations. All the studies share substantially limited ecological validity and artificially disconnected vari-
ables that co-exist in typical interrogations [e.g., minimization and FEPs may occur within the same interrogation
(Kassin et al., 2007) but are generally examined separately in this literature]. The consequences of these limitations
remain unknown, but experts are likely to face challenging cross-examination regarding these aspects of the research,
as recommended by Inbau et al. (2011), among others.

4.4.1 | Participants

The samples employed in the studies are not representative of the sample of individuals charged with criminal acts,
particularly compared with the typical levels of education and the distribution of ages of people in correctional facilities
(Bureau of Justice Statistics, 2003a, 2010), criminal defendants, or police suspects. First, across all studies, traditional
college students are the only participants, almost all of whom appear to come from the social sciences. Second, across
all research methods, the majority of the participants were female, in contrast to the majority of individuals in incarcerated
populations (Bureau of Justice Statistics, 2010) and the majority of drivers stopped by police as well as drivers searched by
police (Bureau of Justice Statistics, 2003b, 2011). The potential limitations of the generalizability of findings as a function
of gender of the participants remain unknown. Additionally, traditional college students may be less vulnerable than
criminal suspects due to college students’ likelihood of being adults rather than children, low likelihood of intoxication,
and low likelihood of cognitive disabilities (Gudjonsson, 2003). Alternatively, college students may be more vulnerable,
particularly in the Alt-key and individual cheating studies, due to their factual innocence (Kassin, 2005, 2012) or to their
likely limited experience with the law, as courts have suggested (e.g., Lynumn v. Illinois, 1963).

4.4.2 | Consequences

The consequences of the perceived transgressions also limit the realism of the research. Scholars working within
ethical limits cannot ask undergraduate participants to risk incarceration, publication of their transgressions to family,
friends, and other communities, and other consequences that come with conviction and incarceration. Researchers
have informed participants that they would face consequences that appear serious to traditional undergraduate
participants (e.g., having to speak to a professor about possible cheating allegations; Russano et al., 2005), and, more
substantially, when Horselenberg et al. (2006) raised the consequences of false confession to paying €250, this
increase in the severity of consequences decreased the false confession rate, but all of these consequences pale in
comparison to actual criminal consequences, particularly for severe crimes.

4.4.3 | Tasks

Perhaps the most serious limitation of the research relates to the transgressions to which the students falsely confess.
Participants in the Alt-key studies confess to negligence, and participants in the social cheating studies confess to
helping a colleague in apparent need, which, as Russano et al. (2005) explained to students who were guilty of
violating test rules, was “an admirable, benevolent, and prosocial act” (p. 484). Selection of these violations reflects
scholars’ ethical goals and requirements for appropriate treatment of participants. Across studies, the transgressions
studied would violate what Elliot Turiel (1983, 2002) and colleagues consider conventional rather than moral rules.
Much research outside the false confession domain has demonstrated that individuals behave differently when a rule
is deemed merely a social convention rather than being based upon moral principles associated with harm and rights.
Perhaps the false confession rate would be much lower if the students, within the constraints of ethical research, were confessing to a moral transgression that harmed another individual rather than to a violation of test norms or of simple arbitrary social conventions.

Additional limits of these conclusions relate to limited realism as necessitated by ethical treatment of participants. As noted previously, researchers cannot ethically require participants to commit actual crimes, expose participants to the actual legal risks of false confession, interrogate participants for hours, or induce the high levels of stress typical in police interrogations (Costanzo & Leo, 2007; Inbau et al., 2011; Jayne & Buckley, 1999; John E. Reid and Associates, 2015b, 2015c; Meissner, Russano, & Narchet, 2010). Scholars have called for increased collaboration with law enforcement officials and other interrogators to increase the realism of research methods while maintaining effective ethical safeguards and to evaluate the diagnosticity of interrogation techniques (Meissner, Hartwig, & Russano, 2010; Woody et al., 2013), and we echo these calls.

Despite the differences between experimental studies and actual interrogations, experimental settings allow for systematic and controlled study of specific techniques and relevant outcomes. Primarily, the factors revealed in the meta-analysis that increase false confessions in experimental studies (e.g., FEPs) also have well-documented consequences for real suspects in the field (Kassin et al., 2010). The advantages of controlled experimental study can provide insights that remain inaccessible in the complex world of actual interrogations.

4.5 Larger implications

The current study demonstrates that this body of experimental work is now large enough for meta-analysis, suggesting that laboratory false confession research is no longer a fledgling field or specialized research paradigm, but rather an established and growing body of tested, reliable, peer-reviewed experimental research accepted by relevant scientific communities that could inform courts (see Costanzo & Leo, 2007; Daubert v. Merrell Dow Pharmaceuticals, 1993; United States v. Hall, 1997). Not only has the body of false confession research demonstrated that innocent individuals falsely confess, but the meta-analysis also extends prior narrative reviews and demonstrates that this is “a research literature that is characterized by eclectic methods that have produced convergent results” (Kassin, 2008, p. 204) in some areas and that some interrogation strategies increase the likelihood of false confession. Experts can readily testify to the outcomes and limitations of the existing experimental studies of false confessions (see Fulero, 2010b; Kassin, 2008) and elaborate upon the effects of moderator variables such as FEPs on confession rates. Although courts vary extensively, as noted previously, we hope that these outcomes can provide experts with increased opportunity to educate courts about police interrogation and the potential for false confessions. Additionally, we hope these meta-analytic findings encourage courts to continue to re-evaluate deception during police interrogation, particularly due to the increased likelihood of false confessions that result from FEPs (see Kassin et al., 2010; Kassin & Gudjonsson, 2004; Woody et al., 2011).

Education of courts and of jurors is paramount due to the prevalence of the myth of psychological interrogation, the belief that innocent people will not falsely confess in the absence of physical coercion or serious mental illness (Leo, 2008; Woody & Forrest, 2009; see also Chojnacki, Cicchini, & White, 2008) and the demonstrated impacts of expert testimony on jurors’ perceptions and trial decisions (Blandon-Gitlin et al., 2011; Gomes et al., 2014; Leo & Liu, 2009; Woody & Forrest, 2009). Beyond jurors, a recent study of sitting judges revealed that judges view expert testimony as relevant for jurors; 72.7% of responding judges reported that they would allow an expert to educate the court about the police interrogation techniques and the possibility of false confessions (Woody et al., unpublished). Additionally, Woody et al. (unpublished) found that judges recognize the deception inherent in FEPs but that judges do not perceive FEPs as coercive; judges, however, were less likely to recommend conviction after reading expert testimony. Experts who present the risks of FEPs to courts have the potential to affect individual judges as well as to shape legal precedents and the larger legal landscape regarding police interrogation. This testimony may become increasingly relevant as courts reconsider whether specific forms of police deception constitute coercion (Bandler, 2014a, 2014b; McKinley, 2014; People v. Thomas, 2014).
Clinical forensic psychologists who serve as experts can also utilize findings from this meta-analysis. As noted previously, most studies in this meta-analysis evaluate college students, most of whom are legal adults rather than juveniles (see Gudjonsson et al., 2006; Redlich & Goodman, 2003; Viljoen, Klaver, & Roesch, 2005), and few of whom have substantial cognitive disabilities (see Gudjonsson & Clare, 1995; Kassin et al., 2010), serious mental illnesses (Redlich, 2004; Redlich, Kulish, & Steadman, 2011) or other factors that increase the vulnerability of a suspect (e.g., high suggestibility; Gudjonsson, 2003). Despite the lack of cognitive and mental health diversity among participants in these studies, these findings demonstrate that some tactics raise the likelihood of confession for participants who are less vulnerable than many actual suspects. If suspects bring additional vulnerabilities into the interrogation room, what impacts might these tactics have above and beyond the impacts on the study participants who are likely to be less vulnerable? Additionally, these tactics may be particularly powerful for those who face multiple vulnerabilities, such as juvenile suspects who struggle with mental illness (Redlich, 2007). Interrogation tactics interact with the identities and potential vulnerabilities of the suspect, and clinical experts can incorporate these findings into their relevant testimony about the larger totality of circumstances.

Due to the limited number of applicable studies as well as the wide variability of methods and moderator variables, we encourage scholars to continue to increase the size of this body of research and to further their commitment to ethical treatment of participants. For this research to continue, ethical standards must continue to evolve with emphases on protection of participants, even with the potentially paradoxical drive for increased realism (e.g., Russano et al., 2005). We join others in calling for increased collaboration with police and other interrogators with emphases on realism, ethics, and diagnosticity (Meissner, Hartwig, et al., 2010, Meissner, Russano, et al. 2010).

We encourage police interrogators and those who train them to consider these findings in their teaching as well as in their practices of interrogation. As Wallace (2010) argued, we perceive interrogators as highly variable practitioners who select methods to fit each suspect, crime, and situation, and we encourage interrogators to consider carefully general approaches as well as specific tactics. In particular, we encourage police interrogators and those who train them to consider carefully the potentially coercive influence of FEPs and other forms of deception during interrogation, particularly for innocent suspects. Fundamentally, we hope that uses and implications of the conclusions from this meta-analysis can decrease the likelihood of false confessions in actual interrogations, which come at an overwhelming price for the accused, the communities in which the crime occurred, and the courts, law enforcement officials, taxpayers, and the criminal justice system as a whole.

ACKNOWLEDGEMENTS

Portions of these data were presented at the American Psychology-Law Society Convention and at the Rocky Mountain Psychological Association Convention. Please send requests for reprints to correspondence author William Douglas Woody. The authors thank student Czarina Grogan for support in the literature searches.

REFERENCES


**How to cite this article:** Stewart JM, Woody WD, Pulos S. The prevalence of false confessions in experimental laboratory simulations: A meta-analysis. *Behav Sci Law*. 2018;36:12–31. https://doi.org/10.1002/bsl.2327