

Development of the Interview and Interrogation Assessment Instrument

by

Jeffrey Kaplan

A thesis submitted to the
School of Graduate and Postdoctoral Studies in partial
fulfillment of the requirements for the degree of
Doctor of Philosophy in Forensic Psychology

Faculty of Social Science and Humanities
University of Ontario Institute of Technology (Ontario Tech University)
Oshawa, Ontario, Canada
April, 2023

THESIS EXAMINATION INFORMATION

Submitted by: **Jeffrey Kaplan**

Doctor of Philosophy in Forensic Psychology

Thesis title: Development of the Interview and Interrogation Assessment Instrument
--

An oral defense of this thesis took place on April 12, 2023, in front of the following examining committee:

Examining Committee:

Chair of Examining Committee	Leigh Harkins, PhD, Ontario Tech University
Research Supervisor	Amy-May Leach, PhD, Ontario Tech University
Examining Committee Member	Joseph Eastwood, PhD, Ontario Tech University
Examining Committee Member	Brian Cutler, PhD, Fielding University
University Examiner	Andrea Slane, PhD, Ontario Tech University
External Examiner	Brent Snook, PhD. Memorial University

ABSTRACT

Confession evidence factors heavily in judicial decision making, and courts may call an expert social scientist to assess the coercive pressures of an interrogation and risk factors for false confession. At present, there exists no standardized methods for performing this task, and each expert uses their own unstructured professional judgment. To address this lack of standardization, we have developed a psychological instrument for evaluating videotaped interrogations: the Interview and Interrogation Assessment Instrument™. We begin with a discussion of the benefits of standardized measurement and proceed to an overview of the conceptualization and initial development of our instrument. In Study 1, we established the bases for the instrument's items and scoring by surveying expert populations. In Study 2 we assessed interrater reliability and explain our instrument refinements based on our results. In Study 3, we examined convergent validity. Social science experts reviewed interrogation videos and rated the coercive pressures along multiple dimensions. We correlated the expert ratings with our instrument's measures. We conclude that the newly developed instrument demonstrates preliminary reliability and convergent validity and appears to be a promising tool for future research and expert consultation in contested confession cases.

Keywords: interrogations, false confessions, expert witness, coercion

AUTHOR'S DECLARATION

I hereby declare that this thesis consists of original work of which I have authored. This is a true copy of the thesis.

I authorize the University of Ontario Institute of Technology (Ontario Tech University) to lend this thesis to other institutions or individuals for the purpose of scholarly research. I further authorize University of Ontario Institute of Technology (Ontario Tech University) to reproduce this thesis by photocopying or by other means, in total or in part, at the request of other institutions or individuals for the purpose of scholarly research. I understand that my thesis will be made electronically available to the public.

The research work in this thesis was performed in compliance with the regulations of Research Ethics Board under REB files # 14173 & 15439

A square box containing a handwritten signature in black ink, which appears to be 'JK'.

Jeffrey Kaplan

STATEMENT OF CONTRIBUTIONS

I hereby certify that I am the first author of this thesis. I have used standard APA referencing practices to acknowledge ideas, research techniques, or other materials that belong to others.

Furthermore, I hereby certify that I am the sole source of the creative works and/or inventive knowledge described in this thesis. This thesis has been published in a peer-reviewed journal as:

Kaplan, J., Cutler, B. L., Leach, A. M, & Eastwood, J. (2022). Development and of the Interview and Interrogation Assessment Instrument. *Psychology, Public Policy, and Law*.

<https://doi.org/10.1037/law0000366>

ACKNOWLEDGEMENTS

This work was supported by the American Psychology – Law Society Grants in Student Aid; Social Sciences and Humanities Research Council of Canada: [Grant Number CGS-D Grant #67-2020-1482]; Social Sciences and Humanities Research Council of Canada: [Grant Number Insight Grant: #430-2016-00605].

I would like to thank Dr. Brian Cutler, Dr. Amy-May Leach, and Dr. Joseph Eastwood for their guidance in refining my research and supervising me through the undertaking of my thesis. This thesis would not have been possible if not for the tireless work of the numerous graduate and undergraduate research assistants who went above and beyond to help in data collection. I further wish to acknowledge Christopher Kelly, Alison Redlich, and Richard Leo for sharing research materials.

TABLE OF CONTENTS

Thesis Examination Information	ii
Abstract	iii
Authors Declaration	iv
Statement of Contributions	v
Acknowledgements	vi
Table of Contents	vii
List of Tables	ix
List of Figures	x
List of Abbreviations and Symbols	xi
Chapter 1 Introduction	1
1.1 Current Assessments and their Limitations	2
1.2 The Interview and Interrogation Assessment Instrument	4
1.2.1 <i>Item Design</i>	6
1.2.2 <i>Rapport Building</i>	7
1.2.3 <i>Evidence-Based</i>	7
1.2.4 <i>Social and Psychological Manipulation</i>	8
1.2.5 <i>Consequence-Based</i>	8
1.2.6 <i>Minimization</i>	8
1.2.7 <i>Investigator Demeanor</i>	9
1.2.8 <i>Suspect Behaviors and Demeanor</i>	9
1.2.9 <i>Physical Environment and Context</i>	9
1.2.10 <i>Suspect Risk Factors</i>	10
1.3 Conceptualization and Development.....	10
1.3.1 <i>Item Selection</i>	12
Chapter 2 Study 1	15
2.1 Participants and Method	16
2.2 Results	17
2.3 Conclusion and Discussion	19
Chapter 3 Study 2	19
3.1 Pilot Study	20
3.2 Method	22
3.2.1 <i>Materials and Procedure</i>	21
3.2.2 <i>Results and Discussion</i>	23
Chapter 4 Study 3	24
4.1 Hypotheses	24
4.2 Method	24
4.2.1 <i>Participants</i>	25
4.2.2 <i>Materials</i>	26
4.2.3 <i>Procedure</i>	27
4.3 Results	29
4.4 Discussion	30
Chapter 5 General Discussion	30
5.1 Limitations	32
5.2 Conclusion	37

References	38
Footnotes	52

LIST OF TABLES

CHAPTER 3

Table 3.1: Pilot study interrater reliability by interrogation	53
Table 3.2: Pilot study interrater reliability by time block	54
Table 3.3: Study 2 interrater reliability by interrogation and time block	55
Table 4: Study 2 interrater reliability for dichotomous measures	56

CHAPTER 4

Table 4.1: IIAI scores descriptive statistics	57
Table 4.2: Study 3 tactic counts and IIAI scores correlated with experts' ratings ...	58

LIST OF FIGURES

CHAPTER 1

Figure 1.1: Active coding of interrogation tactics using the IIAI	59
Figure 1.2: Active coding of suspect behaviors using the IIAI	60

LIST OF ABBREVIATIONS AND SYMBOLS

CJO	Criminal Justice Official
IIAI	Interview and Interrogation Assessment Instrument
ICC	Intraclass Correlation
<i>n</i>	Sample Size
<i>sd</i>	Standard Deviation
SSE	Social Science Expert

Chapter 1. Introduction

All confessions are not equal. Suspects' statements under questioning often play a pivotal role in the prosecution of the guilty. Unfortunately, confessions are also a key contributor to the wrongful conviction of the innocent (Kassin et al., 2010). Although there are no exact statistics on what proportion of confessions are disputed, because of the weight given to confessions among triers of fact confessions are often contested in criminal cases. Concerns about interrogators coercing false confessions from criminal suspects are not modern in origin; those within judicial and social science spheres have been raising objections to certain interrogation practises since the early 20th century (e.g., Borchard, 1932; Munsterberg, 1908; *Wickersham Commission Report*, 1931), if not longer (e.g., *Bram v. United States*, 1897; *Hopt v. Utah*, 1884; *King v. Warickshall*, 1783). Over time, the courts have gradually moved beyond merely examining physical abuse, direct threats, and promises to give more consideration to concepts such as a suspect's "operating mind" and an "atmosphere of oppression" (*R v. Oickle*, 2000, p. 1). This approach offers the courts more latitude in examining confession evidence and creates opportunities for psychological experts to testify on coercion and false confessions.

As judicial interest in the psychology of interrogations has grown, so too has the sophistication of the social science research. The current research base on interrogations and confessions was built upon reliable science and is generally accepted in its field (Gudjonsson, 2021). The body of literature now encompasses experimental research and meta-analyses (e.g., Klaver et al., 2008; Meissner et al., 2012; 2014), surveys of expert populations (e.g., Kassin et al., 2007), observational and archival studies (e.g., Kaplan & Cutler, 2021; King & Snook, 2009; Ofshe & Leo, 1997), and review papers authored by psychologists (e.g., Kassin et al., 2010; Scherr et al., 2020). It has also been highly informed by textbooks and training manuals written

from the perspective of those who conduct suspect questioning (e.g., Carr, 2015; Inbau et al., 2013; Wicklander-Zulawski, 2020), and legal and criminological critical reviews (e.g., Leo, 2008; Pepson & Sharifi 2010). As demonstrated by Kassin et al's (2018) survey of confession experts, there is a strong degree of agreement between experts on risk factors for false confession, although much of this knowledge remains outside of the ken of jurors (Alceste et al., 2021; Jones & Penrod, 2016; Kaplan et al., 2020).

The exponential growth in research over the past few decades has greatly enhanced our understanding of police interrogations, but has also made the prospect of analyzing a given interrogation much more complex; each study that uncovers new knowledge about coercion and false confessions introduces new factors than an expert must be cognizant of and incorporate into their analysis. Important factors may be overlooked and/or influenced by bias. In this article, I introduce a new comprehensive and comparatively objective method of analyzing coercion in interrogations and confessions.

1.1 Current Assessments and their Limitations

In criminal law, expert testimony on confession evidence is typically proffered under one of two contexts. There exist legal protections against being coerced into making statements against one's will under questioning by law enforcement both at the national (*Canadian Charter of Rights and Freedoms*, 1982; *U.S. Constitution, Amendment II*, 1791) and international levels (Melzer, 2018). Criminal court judges may hold suppression hearings to determine the admissibility of a confession by examining its voluntariness (*Miranda v. Arizona*, 1966; Schulhofer, 2001; Thomas & Leo, 2009), and an expert opinion may assist the trier of law in that process. However, absent some explicit violation of a suspect's rights, criminal court judges are typically hesitant to rule a confession inadmissible (Thomas & Leo, 2012). Because of the high

burden of proof in admissibility hearings it is much more likely that the court will consider issues of coercion to go to the weight of the confession evidence, rather than its admissibility, and defer to the trier of fact (Pepson & Sharifi, 2010). In such instances, the jury may be informed about coercion or risk factors for false confessions by an expert psychological witness (Costanzo & Leo, 2007; Marion et al., 2019; Soree, 2005).

In both contexts, an expert would review the interrogation videotape, audiotape, and/or transcript, and idiosyncratically keep note of the factors that they deemed relevant and draw conclusions about the interrogation using their own professional judgment (Kaplan, manuscript in preparation). This is problematic on a few counts. For one, structured and standardized measures usually yield more accurate results and predictions than clinical judgment in most areas of criminal justice-related social science (Grove et al., 2000; Hanson & Morton-Bourgon, 2009; Krauss & Lieberman, 2007).

There are also numerous potential weaknesses and limitations to such an approach, as there is a collection of well-known cognitive biases that may taint evaluators' unguided judgments (see Krauss & Lieberman, 2007; Marion et al., 2019; Neal & Grisso, 2014). Although these biases have not been investigated specifically in interrogation evaluation, they are well documented in other areas of forensic assessment. In the field of risk assessment, for instance, unguided judgments have been repeatedly shown to be vulnerable to influence from the *representativeness heuristic* (Krauss & Sales, 2001; Monahan, 1981), which is to judge something/someone by how closely it matches a schema (Tversky & Kahneman, 1974; 1986). In the context of interrogation evaluation, this could hypothetically manifest as evaluations being subtly and subconsciously influenced by their own preconceptions of how coercive interrogations unfold or how vulnerable suspects behave.

Other vulnerabilities may be due to the limitations of attention and memory, and related phenomena such as *selective attention* (Watts & Weems, 2006) and the *availability heuristic* (Tversky & Kahneman, 1973). When experts use unstructured methods, some relevant factors may escape their attention or not be given proper consideration; an effect which has been noted in medicine, mechanics, aviation, and many other fields (Gawande, 2010).

Other biases may arise due to the nature of North American legal systems. Experts' assessments of risk of recidivism and civil commitment criteria tend to show a systematic bias, or *adversarial allegiance* (Murie et al., 2009), towards the determinations preferred by the party that retained them (McAuliff & Arter, 2016; Rufino et al., 2012), with moderate to large effect sizes (Murrie et al., 2013). Subjective unguided measures are more susceptible to practitioner bias and adversarial allegiance than highly structured methods of assessment (Gowensmith et al., 2010; Murrie & Boccaccini, 2015).

A final limitation of evaluating interrogations and confessions in the absence of any structured instrument is that the resultant assessments lack any firm metrics. If an expert were to be asked how an interrogation and confession compare to others (a question that, as a consultant, I am routinely asked by defense lawyers who send me recorded interrogations for review), the only answer they could proffer would be a qualitative statement based on others they have reviewed in their careers. Thus, the process of evaluating interrogations is also vulnerable to biases resultant from an *anchoring effect* (Tversky & Kahneman, 1973) wherein experts may, in part, base their opinions on implicit comparisons to other interrogations they have encountered rather than the collective knowledge of the social science on interrogations and confessions.

1.2 The Interview and Interrogation Assessment Instrument™

To address the above-mentioned limitations and to introduce standardization to the evaluation of interrogations and confessions, I undertook the development of the *Interview and Interrogation Assessment Instrument*TM (IIAITM)¹. The IIAITM was designed to capture factors that affect the coercive pressure of suspect questioning, and thereby serve as a measure of the risk of a coerced compliant false confessions. These include factors that relate to the individual being interrogated, those that relate to the environment of the interrogation, and the tactics used during the interrogation. The recording of interrogations is standard practice in many jurisdictions (Bang et al., 2018), and the framework I developed is designed to function as an observational coding structure for use while viewing videotaped interrogations (see Figures 1 and 2).

Although the primary focus of the instrument's measures pertain to factors in accusatory interrogations, I opted to include the word "interview" in the name of my instrument for several reasons. Many models of suspect questioning such as The Reid Technique of Interrogation (Inbau et al., 2013) and the Phased Interview Model (Carr, 2015) begin with an interview which may shift to an interrogation; knowing what took place in the pre-interrogation interview is highly pertinent to evaluating the proceeding interrogation. To test and validate the IIAITM I primarily used videotapes of interrogations, although many included a pre-interrogation interview as well. Secondly, if non-accusatory and non-coercive methods are used to gain a confession, without the use of any accusatory tactics, my instrument may then be used to demonstrate that a confession was obtained without putting the suspect under significant coercive pressures.

It should be made explicitly clear that the IIAITM was designed to measure coercion in interrogations and not the absolute truth of confessions. A coerced confession may be a true

confession (i.e., when the suspect is guilty) or a false confession (i.e., when the suspect is innocent). False confessions also come in various forms. A *voluntary* false confession is one in which an innocent individual offers a confession with little or no prompting (Kassin et al., 2010). A *coerced compliant* false confession occurs when an innocent suspect is persuaded or coerced by investigators to confess (Kassin et al., 2010). In this research I focus specifically on assessing coercion in the interrogation process. I therefore focus on risk factors for coerced compliant false confessions, but the IIAI™ was not designed to assess the veracity of confessions.

The IIAI™ does not replace an expert witness. It provides a means of organizing findings and produces metrics related to an interrogation and confession. Interpreting that data requires a strong background in the psychology of interrogations. Thus, the IIAI™ is meant to aid those who already possess a mastery of the psychology of interrogations and confessions in evaluating confession evidence and organizing their opinions. IIAI™ evaluations contain a record of what occurred in an interrogation, but the psychology underlying why the interrogation may have been coercive or why the identified occurrences are risk factors for false confession still requires the interpretation of a qualified expert. The IIAI™ can help an expert keep track of all of the pertinent factors and produce metrics about an interrogation that no other methods of evaluation currently offer. Making sense of those metrics and explaining them to the courts is the responsibility of the expert witness.

1.2.1 Items Design

The IIAI™ items are divided into tactics, the suspect's behaviors and demeanor, investigator demeanors, the physical context and environment, and suspect risk factors. The tactic items of the IIAI™, as I chose to define them, generally do not refer to extremely specific behaviors, but rather broader constructs. They may, therefore, manifest in an interrogation in

various ways. For instance, the item *Degrading Others* refers to an interrogator placing blame on the victim or a co-accused, whereas *Justifications* refers to blaming something more abstract, such as a temporary loss of control, desperation, or a black out theme. The reason for this design is that interrogations can vary from one another in a virtually infinite number of ways. The details of the crimes, legal and social contexts, characteristics of the suspects, victims, and investigators may differ from one case to another across countless dimensions. Using broader heuristic categories, rather than fixed sets of specific behaviors, allows the IIAI™ the flexibility needed to be applicable to a wide variety of investigations. Each tactic, in turn, is organized into the following domains modelled after those first identified by Kelly et al. (2013).

1.2.2 Rapport Building

Rapport Building contains tactics aimed at establishing a social connection between the investigator and suspect, which is most likely to occur in the early stages of an interrogation (Kaplan & Cutler, 2021; Kelly et al., 2016). Engaging in rapport building has been demonstrated to increase cooperation and the amount of information gathered from criminal suspects (Gabbert et al., 2021), as well as possibly increasing susceptibility to misleading information (Vallano & Schreiber Compo, 2015). Some commentators have argued that the establishment of rapport during the adversarial context of a suspect interrogation is inherently deceptive and is independent of, and often in opposition to, the suspect's best interests (Crough et al., 2022). The items under this domain are engaging in unrelated conversations and empathizing/sympathizing with suspects. The domain also includes the more coercive items of the investigator making statements that they are trying to help the suspect and offering incentives for information.

1.2.3 Evidence-Based

The *Evidence-Based* domain contains tactics that leverage evidence or perceptions of proof over the suspect in some manner. It includes directly presenting evidence, pointing out contradictions in a suspect's account, as well as deceptive tactics, such as presenting false evidence, bluffs/bait questions, and referring to the strength of the evidence without revealing specifics.

1.2.4 Social and Psychological Manipulation

During an accusatory interrogation, an investigator may employ certain *maximization techniques* (Kassin & McNall, 1991) that emphasize the potential consequences the suspect is facing (Inbau et al., 2013; Kassin et al., 2010). Typically, these consequences in some way refer to charges or sentencing; however, on occasion, the negative outcomes that an investigator chooses to emphasize are those related to social outcomes, psychological suffering, or even religion. The *Social and Psychological Manipulation* domain contains such tactics, including increasing feelings of guilt, attacking a suspect's character, and demeaning comments.

1.2.5 Consequence-Based

The majority of the maximization techniques accounted for by the IIAI™ are found under the *Consequence-Based* domain, which contains tactics that create punishments for non-cooperation and emphasize the harms of further non-cooperation. The domain includes items such as directly accusing the suspect, making direct threats, expressing certainty of conviction and harsh sentencing, emphasizing the seriousness of the crime, and directly presenting a forced choice between two alternative themes.

1.2.6 Minimization

The items comprising the *Minimization* domain all imply mitigated punishment in some way. It includes offering justifications, appealing to suspects' self-interest, and downplaying the

seriousness of the offence. These items can range from the investigator making simple statements to constructing sophisticated themes which allow the suspect to excuse, normalize, and/or rationalize their criminal actions. The domain also includes direct promises of leniency, a tactic prohibited in most jurisdictions.

1.2.7 Investigator Demeanor

Aside from the five tactic domains listed above, several other types of variables are accounted for by the IIAI™. The *Investigator Demeanor* domain contains a set of items that are coded in a dichotomous format directly on the IIAI™ scoring sheet. These are judgments about investigator's overall behavior throughout the interrogation, such as their use of *Aggressive Questioning* (e.g., highly accusatory tone, asking questions or employing tactics a rapid-fire manner where the suspect is not given the opportunity to respond).

1.2.8 Suspect Behaviors and Demeanor

There are a small number of suspect behaviors pertinent to the evaluation of the interrogation that are accounted for by the IIAI™. These include behaviors indicative of distress (e.g., crying), indicative of resistance (e.g., *Denials* and *Objections*), making inculpatory statements, and legally relevant behaviors (e.g., requesting an attorney). The IIAI™ also keeps track of the suspect's general demeanor. One of five demeanors are possible. These demeanors are conglomerations and amalgamations of more specific behaviors that the content experts who were consulted in the development of the IIAI™ indicated were pertinent to evaluating a suspect's level of emotional duress (see Figure 2).

1.2.9 Physical Environment and Context

Physical Environment and Context is accounted for on the scoring sheet and includes factors such as isolation before interrogation, holding the interrogation in a small and barren

room, and the use of a polygraph. It also includes a Y/N item to account for whether the suspect was read a proper legal caution at the start of questioning.

1.2.10 Suspect Risk Factors

There are nine *Suspect Risk Factors* accounted for in checklist form on the IIAI™ sheet. These include intellectual disability, mental illness, sleep deprivation, and intoxication. It is not at the IIAI™ user's discretion to make these determinations based solely on what they observed in the videotaped interrogation. Rather, this information is generally found in the case's discovery package in the form of police reports, hospitalization records, and evaluations conducted by psychiatrists and clinical psychologists.

1.3 Conceptualization and Development

The above represents the result of six years of research and three separate studies. To explain how I arrived at this current point, I begin with a basic summary of the early concept development and literature review. Before one can endeavor to measure a construct, that construct needs some tangible operational definition. Following a thorough review of the legal and psychological literature, I arrived at an operational definition of coercion in interrogation on which to base the instrument's measures: "Coercion in police interrogation consists of the use of persuasive techniques that limit the suspect's autonomy by manipulating the perceived costs and benefits of possible courses of action and/or depleting the suspect's motivation or ability to resist persuasion and acceding to the investigators' demands" (Kaplan et al., 2019, p. 6).

Part of this definition was borrowed from Leo and Liu (2009), who had defined coercion as "[that which] removes an individual's perception of their freedom to make a meaningful choice during a police interrogation" (p. 385). Once a suspect waives their right to silence and an interrogation begins, they essentially have the two options of denial or confession; many

interrogations are premised on making confession seem the better option. Coercive pressure may be applied by manipulating the perceived costs and benefits of confessing or continuing with denials, and thereby limit the suspect's perceived viable options. Our operational definition therefore includes *choice architecture* (manipulating perceived options and their pros and cons; Knowles & Linn, 2004). The concept of imposing constraints on choices and contrasting them against conditional offers was also, coincidentally, the basis of Luna's (2022) recent definition of coercion in interrogations and plea bargaining. Luna (2022) defined coercion as arising when "*P* (i.e., the coercer) is in a position of power and communicates a conditional offer and/or threat to *Q* (i.e., the coerced) who is in a constrained situation created by *P*" (p. 7). Essentially the investigator constrains the suspect's options by removing denial as a viable course of action either by making it seem impossible or tying negative consequences to it, and then makes a conditional offer that indicates that confessing will lead to more desirable outcomes.

My definition of coercion, however, differs from that offered by Luna (2022) in two respects; it does not explicitly address the power differential between suspect and investigator (an insightful connection worthy of consideration going forward), and it includes the psychological wearing down of a suspect via *interrogation-related regulatory decline* (Davis & Leo, 2012) that is absent from Luna's (2022) definition. The ability to resist the demands of others requires some powers of concentration and self-regulation (Knowles & Rinier, 2007; Muraven & Baumeister, 2000), which an investigator may aim to deplete (Davis & Leo, 2012). A suspect's ability to critically evaluate information, weigh their options, and resist an investigator's demands may be diminished through repeated coercive tactics that cause fatigue and emotional duress (Evans et al., 2014). Our cognitive resources can be metaphorically thought of as an energy source, which can be drained. Eventually thinking and reasoning

processes become seriously impaired, the ability to withstand any more stress dissolves, and the suspect gives in to the demands of the investigator to confess (Davis & Leo, 2012; 2013).

It must be noted that my use of the term coercion differs from legal definitions which center on (in)voluntariness. When a confession is given, it is viewed by the legal system in dichotomous terms of voluntary or involuntary, where in the latter no option existed but for a suspect to capitulate to the demands of the criminal justice system (*Hopt v. Utah*, 1884; Philips, 1984). Coercion, by my definition, is instead view on a continuum, and I do not identify an inflexion point at which interrogations cross a definitive boundary into being coercive.

My operational definition provided a benchmark to gauge the types of variables that would need to be measured in order to effectively evaluate interrogations. Identifying observable and measurable coercive occurrences was necessary to form the foundations of an assessment instrument for evaluating interrogations. The selection of items, in turn, necessitated the above operationalization of the term coercion to serve as the criterion by which potential items were selected. This process was cyclical. Due to its intended applications, in constructing the IIAI™ and selecting potential items for inclusion, special attention was also paid to those risk factors that have been empirically connected to coerced compliant false confessions such as presenting false evidence (Klaver et al., 2008).

1.3.1 Item Selection

For the purposes of interrogation assessment, I began mining the literature for the specific interrogation tactics that could contribute to, or potentially detract from, the coerciveness of an interrogation. Some of the first sources consulted were those that provided overviews of interrogations, confessions, and risk factors for false confession such as review papers (e.g., Kassin et al., 2010) and books on interrogation (e.g., Gudjonsson, 2003; Leo, 2008).

Some of my primary sources of tactic items were past observational studies; because these items had been used previously in the evaluation of interrogations, accounting for them within my own observational framework was considered achievable. I retained the same phrasing from the original sources wherever possible. Leo (1996) and King and Snook (2009) provided a number of items representing interrogation tactics, most of them related to the Reid Technique of Interrogation (Inbau et al., 2013). Other items were drawn from observational studies conducted by Pearse and Gudjonsson (1999), Bull and Soukara (2010), and Kelly et al. (2016). Items were also drawn from relevant case law (e.g., *Malinski v. New York*, 1945). A number of surveys regarding interrogation and interrogation tactics were consulted in order to identify coercive tactics for possible inclusion in the IIAI™. One such source was a survey conducted by Kassin et al. (2007), which, among other things, asked police officers about the tactics they regularly use. Expressing anger at suspects, bluffing about evidence, and leveraging polygraph results as evidence during an interrogation were items added to the IIAI™ from this survey. Other surveys that I reviewed included those that asked laypeople how they perceive coercion in interrogations (e.g., Henkel, et al., 2008), and one that surveyed forensic psychological experts (Kassin et al., 2018).

It is difficult to deem an interrogation as being coercive without considering the effect on the target of the coercive action (Lidz et al., 1997; Nozick, 1969). Items representing suspect behaviors indicative of being under coercive duress were gathered from observational studies such as King and Snook (2009) and Kelly et al. (2015), and a meta-analysis by DePaulo et al. (2003). Given the intended application of the instrument, other legally relevant suspect behaviors (e.g., requesting an attorney) were also included among the instrument's measures.

Another set of variables I endeavored to incorporate into the measurement of coercion was the environment and context of the interrogation. The length of police questioning was of particular concern, considering that known false confessions in the past have often been the product of excessively long interrogations (Blair, 2005; Drizin & Leo, 2004; Leo & Ofshe, 1998). As to the physical setting itself, a very small interrogation room where the investigator is standing in between the suspect and the exit may create an atmosphere that amplifies the coercive nature of the situation (Trainum, 2016). There was also a significant amount of literature on the effects of isolating suspects before and during interrogations. Isolating suspects before an interrogation is common practice (Inbau et al., 2013), which is meant to make criminal suspects feel cut off from social supports. This may decrease suspects' sense of autonomy and alter the power dynamics of the situation in a manner that makes it easier for investigators to cut off possible courses of action and manipulate what the suspect perceives to be their options in the interrogation (Scherr et al., 2020). Based on literature review of risk factors for false confession items such as length, isolation, the size of the room, and the time of day were added to the IIAI™ under the category of Physical Environment and Context.

Personal characteristics that enhance vulnerability to coercion or the risk of false confession were also relevant items. Suspect Risk Factors, as defined in the IIAI™, are permanent or semi-permanent traits held by suspects that increase their susceptibility to coercion and risk of coerced compliant false confessions. Mental illnesses and intellectual disabilities are risk factors for false confession that have been demonstrated through the examination of past false confessions (e.g., Redlich, 2004; 2007; Redlich et al., 2011; Schatz, 2018), and discussed at length theoretically (Gudjonsson, 2003; 2010). There has also been a great amount of research on youth as a risk factor for false confession (e.g., Cleary, 2017; Malloy et al., 2014; Redlich,

2009). Other items added to this section of the IIAI™ were temporary states such as sleep deprivation (Blagrove, 1996; Frenda et al., 2016).

A systematic review performed by Kelly et al. (2013) was my greatest resource throughout the entire development process. The authors set out to create a taxonomy of interrogation tactics; to this end, they compiled a list of 800 interrogation tactics, combined the redundant tactics, removed those that were not applicable, and divided the remaining tactics into categories they referred to as *domains*. Although I have made considerable alterations to the taxonomy as originally envisioned by Kelly et al. (2013), such as moving all of the minimization items under one domain and renaming the domains, the general structure of mesolevel domains that they identified has been retained in the IIAI™.

At the end of my literature review, we identified 194 separate items (98 tactics across five domains, 65 suspect behaviors, 18 environment and context items, and 13 suspect risk factors) for potential inclusion in the IIAI™. I then used item reduction strategies (Study 1) and reliability analyses (Study 2) to settle on my final set of items and scores for my validity tests (Study 3).

Chapter 2. Study 1

Study 1 focused on item reduction and had four research objectives. The first was to drop any of the 194 items that lacked coercive properties; it would be unwieldy for the user to try and account for such a large number, and I suspected that many of the items were not relevant to an instrument measuring coercion. Second, I sought to further reduce the number of items by combining those items that were conceptually similar and that experts thought were comparable. Third, because some of the items may have been vaguely phrased or were operationally unclear, I wanted experts to indicate which items might cause confusion for future IIAI™ users or be too

ambiguous to judge. Fourth, I wished to establish that the remaining items did, indeed, have coercive properties and to estimate the relative coercive magnitude of each item.

2.1 Participants and Method

I chose to survey two different groups of experts. The first group were *social science experts* (SSEs) specializing in interrogations and confessions. The recruitment letter for SSEs in Study 1 requested their participation if they had “research interests and experience in police interviewing of suspects and witnesses, coercion, or false confessions.” The reasoning for using their expert knowledge as an empirically based assessment of each item was that they would be highly familiar with the research literature, publish and contribute to that literature, and that many of them would have applied experience in evaluating interrogations as expert witnesses. Thus, they would have the expertise needed to judge the coerciveness of the list of items I had compiled. *Criminal justice officials* (CJOs) from North America also completed the survey and were invited to a semi-structured interview. CJO participants were recruited if they had experience “conducting forensic interrogations or evaluating confession evidence as an officer of the court.” Thus, the CJO sample was comprised of police investigators, defense attorneys, and prosecuting attorneys. This provided a sample of experts from a variety of perspectives.

A total of 74 participants completed the survey: 54 SSEs and 20 CJOs. The average age of participants was 45.77 ($SD = 13.94$). The sample was 37.8% ($n = 28$) male and 62.2% ($n = 46$) female. The participants were 84% ($n = 62$) White, and 16% ($n = 12$) identified as another race or as biracial. For most items, the ratings of SSEs and CJOs did not statistically differ from one another (see Kaplan et al., 2020).

The survey was hosted by Qualtrics. It contained all 194 items identified through literature review, which were divided into four categories: tactics, suspect behaviors,

environment and context, and suspect risk factors. Each item was rated on a 7-point Likert scale. Each scale also had an additional option to indicate an item was too vague or unclear to judge. The text above each item included our operational definition, and asked participants "...how coercive would you rate the following occurrences?" Participants rated each interrogation tactic on a seven-point Likert scale ranging from -3 (*Coercion reducing*) to 3 (*Coercion increasing*), with 0 marked "Neutral." We included negative integers because there was some indication from past research that certain behaviors may reduce feelings of coercion (specifically, those that increase feelings of independence and efficacy; Norvoll et al., 2017), such as allowing the suspect to control the conversation. Thus, we expected that certain items, such as allowing the suspect to control the conversation, may be perceived by experts as detracting from feelings of duress and an atmosphere of coercion. The scales for the suspect risk factors slightly differed. The question text above these scales read: "Given your own personal understanding of coercion in police interviews and interrogations, to what degree do the following characteristics increase a suspect's susceptibility to coercion?" The scale ranged from 1 (*Not at all*) to 7 (*To a large extent*), again with the option "Vague or Unclear." This survey was administered as part of Kaplan et al. (2020), which was a secondary use of this data. This study was not preregistered. The data associated with this study is available through the Center for Open Science at <https://osf.io/r27uq>. This study was approved by my institution's Research Ethics Board.

2.2 Results

All survey items were first subjected to one sample *t* tests. The interrogation tactics, suspect behaviors, and environmental considerations were compared to the neutral score of zero; suspect characteristics were compared to the lowest score of one. Those that did not significantly differ were removed. I did not adjust the alpha levels for these tests because testing

was not disjunctive where one significant result would be needed to reject a common null hypothesis. Because the study was not meant to explore a joint hypothesis, but rather each test had its own null hypothesis, I believed that adjusting the alpha level to account for 194 tests to be inappropriate and that it would greatly inflate the chances of Type II errors. As well, when five percent or more of participants labelled an item as vague or unclear, the item was discussed by my research team and removed if the team agreed that the item was too ambiguous to judge from a videorecorded interrogation.

I then combined items that were conceptually similar and were given similar ratings by the experts that I surveyed. Two items relating to the use of a polygraph were also combined, despite having a lower Cronbach's alpha (see online materials: <https://osf.io/r27uq>). Ultimately, my analyses of survey results reduced the number of interrogation tactics items from 98 to 51. Using Kelly et al's (2013) taxonomy as my guide, I then organized those tactics into the domains. Those tactics that aimed to create a social connection between suspect and investigator were put into the Rapport Building domain ($n = 8$ items; *Cronbach's* $\alpha = .614$). Tactics that leveraged perceptions of proof were organized into the Evidence-Based domain ($n = 11$ items; *Cronbach's* $\alpha = .762$). Tactics that leveraged social and psychological outcomes were organized into the Social and Psychological Manipulation domain ($n = 10$ items; *Cronbach's* $\alpha = .740$). Those tactics that created or emphasized punishments for non-cooperation were put into the Consequence-Based domain ($n = 17$ items; *Cronbach's* $\alpha = .813$). Finally, tactics that implied mitigated punishment were placed into the Minimization domain ($n = 5$ items; *Cronbach's* $\alpha = .803$).

I reduced the suspect behaviors from 65 to ten by first removing eight suspect behaviors due to having neutral properties, and combining two that were similar. I then grouped the

remaining suspect behavior items into *Suspect Demeanors* both conceptually and by the ratings given by experts. Based on those verbal and physical behaviors that grouped together, I identified five general suspect demeanors (*Fearful, Anxious/Agitated, Withdrawn/Defeated/Depressed, Calm/Neutral, and Defiant/Belligerent*). Removed and combined items from Study 1 and their associated statistics can be found in my online materials (<https://osf.io/r27uq>).

2.3 Conclusion and Discussion

After analyzing the data from our survey, I had an organized collection of items that related to my operational definition of coercion and had been established by experts as having coercive properties. The survey results indicated that the items were internally reliable, what remained was to establish interrater reliability by testing whether they could be reliably accounted for by users. In my next study, I collected videos of criminal interviews and interrogations conducted in the US and Canada, and recruited research assistants to evaluate them using drafts of the IIAI™.

Chapter 3. Study 2

The IIAI™ is powered by V-Note™ software, shown in Figures 1 and 2. Each tactic and suspect behavior item of the IIAI™ has a key that is used to designate that the corresponding item is occurring. V-Note™ keeps track of which items were coded, and when, on its timeline. This data can be exported into spreadsheet format with counts of tactics and suspect behaviors, the length of time they were occurring, when in the video they were occurring, and with which other tactics and behaviors they closely co-occurred. Environment and Context items and Suspect Characteristics are recorded manually in “Y/N” format on the IIAI™ sheet. In the upper right corner, there is a space available for making notes during the assessment of the

interrogation, such as the type of evidence raised or the types of justifications offered (see Figure 1). The timelines from the tactic and suspect behavior coding can then be combined and compared, which improves the efficiency of contamination analyses by making it clear who brought up what piece of information first.

3.1 Pilot Study

Undergraduate research assistants were given a brief tutorial on V-Note™ software, provided with the IIAI™ coding legend, and then used the IIAI™ to code interrogation videos along with the first author. This process entailed independently reviewing and coding interrogations using the IIAI™, then meeting after coding ten interrogations each to compare our results and assess interrater reliability. Areas in which coding differed between raters were discussed, steps were taken to resolve any ambiguities (such as rephrasing items), and I made continual adjustments based on our experiences applying the instrument. This process took place over approximately 18 months and involved the coding of 50 interrogation videos (i.e., five sets of ten interrogations).

I used two general units of measurement when testing interrater reliability: (1) interrogation and (2) time block. For instance, if one user coded 39 pieces of evidence being presented for a given interrogation and another coded 31, those two data cases would become a point of comparison for analysis using each interrogation as the unit of measure in intraclass correlational (ICC 2, k) analyses. Those measures, however, would not necessarily mean that both users had coded the same occurrences. When the exported data from two raters are compiled together, it forms a cross-tab matrix of tactics and 90-second time blocks (each column a rater's coding, each row a 90-second time block). Intraclass correlational (ICC 2, k) analyses were performed on both the 90-second time block data and the per-interrogation data to assess

the reliability of our coding. ICC coefficients may be interpreted using the following guidelines: less than 0.40 = poor; between 0.40 - 0.59 = fair; between 0.60 - 0.74 = good; between 0.75 and 1.00 = excellent (Cicchetti, 1994). Suspect risk factors and environment and context items were analyzed using Cohen's κ .

The reliability measures from the first rounds of pilot testing ranged from moderate for the per-interrogation data (ICC 2, $k = .556$) to poor for the time block data (ICC 2, $k = .384$). One of the primary reasons for low interrater reliability were that there were too many items to attend to at once. To address this, the number of tactic items was reduced from 51 to 36 by dropping extremely low frequency items. Any item that occurred less than five times during testing was reviewed by the research team and dropped if we agreed that they were not necessary for measuring coercion in police interrogations. For instance, the items *Eye Contact*, *Unexpected Question*, and *Reverse Order* were three tactics taken from Vrij et al.'s (2006) overview of detecting deception through increasing cognitive load. These tactics did not occur in any of the interrogations during pilot testing, and had questionable coercive properties, and thus these items were removed. There were, however, some very rarely occurring items that were retained due to their clearly coercive properties, such as the investigator becoming physically aggressive or telling the suspect they cannot leave until they confess. It would detract from the instrument to too great a degree if it could not account for such occurrences, despite their very low prevalence.

I also combined items that were being mistaken for each other, had comparable coercion ratings, and shared conceptual similarities. For example, in pilot testing there was low agreement for the items of *Increase Guilt* and *Honor/Religion*. A great deal of the responsibility for the lack of interrater agreement lay in the fact that in places where one rater had coded one of these items, a second rater would code the other. These items had been given similar coercion

ratings by the experts surveyed and all went to the same general construct. Therefore, for the sake of decreasing the number of items, these were combined under Increase Guilt. Items that were removed or combined during pilot testing can be found in our online materials:

<https://osf.io/r27uq>.

Based on item clarification, item removals, and item combinations over multiple rounds of coding, reliability steadily improved to the point that it could be considered good-to-excellent. Summarized results can be found in Tables 3.1 and 3.2. I then endeavored to formally establish reliability in the following study.

3.2 Method

3.2.1 Materials and Procedure

Previous research assistants who had used the IIAI™ were junior undergraduates with little education on interrogations and confessions who were given relatively little training. Due to their difficulties in understanding some of the more nuanced aspects of interrogations, I compiled a 78-page training manual and a training program. I recruited three new research assistants holding master's degrees who were currently completing doctoral studies, and two new research assistants in their third year of bachelor studies in psychology, but not yet holding any degrees or certifications, to undergo training.

The training program took place over two days. During the first day of the course, an 85-slide presentation was given to the research assistants; it took approximately 5 hours to complete. This first session gave an overview of coercion in police interrogations, and covered the Suspect Behaviors, Environment and Context items, and Suspect Risk Factors. The second part of the course covered the coding of interrogation tactics and Investigator Demeanors. It included a 121-slide presentation and took approximately 6 hours to complete. Following the completion of

training, the five research assistants were given seven interrogation videos totalling 6 hours and 55 minutes to code, which we then analyzed for interrater reliability. This study was not preregistered.

3.2.2 Results and Discussion

I established acceptable levels of interrater reliability for most items and domains, with the exception of Rapport Building and Suspect Behaviors (see Table 3.3). For Rapport Building, this was often due to disagreements regarding coding rapport building conversations as one long instance or as many instances. At times, the research assistants also coded very basic procedural questions (e.g., name, date of birth, etc.) as rapport building conversation, when investigators were merely filling out necessary paperwork. Similar issues of coding one long instance or multiple instances arose when coding the suspect's behaviors, as well. Disagreements also arose when research assistants coded denials and admissions to acts that were not criminal offenses, such as failure to pay bills or having extramarital affairs. Reliability for Social and Psychological Manipulation and Consequence-Based tactics were also weaker than expected. Again, this was largely due to whether a tactic was coded as one long instance or multiple instances. Some specific items such as *Evidence Bait*, *Downplaying Seriousness*, and *Alternate Question* were recurring sources of disagreement due to being under coded by some research assistants but attended to by others. Interrater reliability statistics on each individual item are available through our online materials (<https://osf.io/r27uq>).

The total interrater reliability of the instrument was excellent for both the per-interrogation data (ICC 2, $k = .944$) and the time block data (ICC 2, $k = .856$), and research assistants agreed about how they captured the main tactics of accusatory interrogations. ICC measures for the Evidence-Based domain in particular were excellent for the per-interrogation

data (ICC 2, $k = .970$) and the time block data (ICC 2, $k = .925$). There was also little disagreement between research assistants for the dichotomously recorded items of the Environment and Context and Suspect Risk Factors. Other specific domains and tactics showed poor interrater reliability, and will be the focus of future revisions and additional interrater reliability research.

Chapter 4. Study 3

Each item was given a weighted coercion score based on an average of the experts' ratings from Study 1. As tactics are used and recorded, the total tactic score and domain scores continues to increase relative to those tactics' coercion weight. I initiated Study 3 to examine and establish the validity of those scores. To accomplish this, I asked 36 social science experts (SSEs) to review and evaluate one videotaped interrogation each. All interrogations had been previously assessed using our instrument, and SSEs were blind to these previously calculated scores.

Experts rated the coercive pressures of the interrogation in total and along dimensions reflecting the domains of the IIAITM. Their ratings were compared to IIAITM domain scores, suspect behavior scores, and total IIAITM scores. The purpose of these comparisons was to demonstrate the convergent validity of the IIAITM with the professional judgment of social science experts currently working in the field of interrogation and confession evaluation.

4.1 Hypotheses

The first hypothesis was that IIAITM tactic measures would positively correlate with independent SSEs' evaluations of the total coercive pressure of the interrogations. The second hypothesis was that the experts' ratings for specific types of coercive pressure (e.g., Evidence-Based, Minimization etc.) would positively correlate with those measures from the

corresponding IIAI™ domains. The third hypothesis was that SSEs' ratings of suspect duress would correlate positively with IIAI™ measures of suspect behaviors.

4.2 Method

4.2.1 Participants

Because I expected a moderate correlation between SSEs' ratings and instrument scores ($\rho = .55$), power analysis conducted in G*Power 3.1 (Faul et al., 2007) indicated that a sample of 35 SSEs would be sufficient to statistically demonstrate the relationship. This was rounded up to 40, but I was ultimately only able to collect 36 responses from qualified SSEs.

Participants were recruited through the American Psychology - Law Society listserve, from personal contacts, and by referral from other participants. The inclusion criteria for SSEs were 1) that they held a graduate degree in a relevant area of the social sciences, 2) they had performed expert witness functions pertaining to assessing interrogations (e.g., testifying in court, writing reports or affidavits) and/or had published peer-reviewed research in the areas of interrogations and confessions, and 3) were based in North America at the time of the study, due to the fact that the IIAI™ was developed for use in North American legal systems. Each SSE was compensated with a \$100 CAD or \$75 USD Amazon gift card for their participation. This study was approved by my institution's Research Ethics Board.

The 36 participants were 52.8% female ($n = 19$), 44.4% male ($n = 16$), and 2.8% ($n = 1$) declined to identify their gender. Participants ranged from 25 to 73 years of age ($M = 42.58$, $SD = 13.83$). In total, 86.1% ($n = 31$) of SSEs were White, 8.3% ($n = 3$) were multiracial, and 5.6% ($n = 2$) were Black. Of the 36 SSEs, 41.7% ($n = 15$) reported never serving as an expert witness but having published peer reviewed research, 36.1% ($n = 13$) reported serving in one to ten

criminal cases of disputed confessions, 5.6% ($n = 2$) reported serving in 11-24 cases, 11.1% ($n = 4$) reported serving in 25-49 cases, and 5.6% ($n = 2$) reported serving in 50 or more cases.

To ensure that participants had completed the study as required, I took note of when each participant began and completed the study. If the total amount of time was less than the length of the interrogation video that they were assigned, their responses would have been removed from the data set as it was not possible for them to have completed watching the video before filling out the evaluation. As an attention check, participants were asked “Did the suspect confess?” No SSEs were removed for suspiciously low participation times or failing the attention checks, nor did any begin the study and not complete participation.

4.2.2 Materials

The study was hosted by Qualtrics, and participation took place completely online. Qualtrics randomly assigned each participant to a video by providing a link to a Google Drive folder that contained a videotaped interrogation and a brief description of the case. After reviewing those materials, participants returned to Qualtrics and completed their evaluations via that online platform.

Interrogation Videos. The 36 videos used in this study ranged from 26-138 minutes ($M = 71.85$, $SD = 26.3$) and were of interrogations that took place in Canada or the United States between 2003 and 2018. Interrogations were chosen if they took place in North America, the suspects were 18 years of age or older, and the videos contained the interrogation in its entirety. The suspects in the videos ranged in age from 18 to 71 ($M = 37.0$, $SD = 12.89$); 89% of them were male ($n = 32$) and 11% ($n = 4$) were female. The majority of the suspects (83.3%; $n = 30$) were White, 13.9% ($n = 5$) were Black, and 2.8% ($n = 1$) were Latinx. All interrogations pertained to offences that could be charged at the indictable/felony level: 52.8% ($n = 19$)

homicide, 13.9% ($n = 5$) grand theft, grand larceny, or fraud, 13.9% ($n = 5$) possession of child pornography and/or soliciting minors for sex, 5.6% ($n = 2$) sexual assault or rape, 5.6% ($n = 2$) assault and battery or attempted homicide, 5.6% ($n = 2$) reckless disregard for life and fleeing police, and 2.5% ($n = 1$) weapons possession in relation to terrorism. Two of the defendants who confessed were later found not guilty at trial. Twenty-one of the videos had been previously used as research materials in Kaplan and Cutler (2021); otherwise, to my knowledge, none of the videos had been used in any previous studies.

4.2.3 Procedure

At the time of contact each participant was presented with a recruitment letter explaining the purpose and the goals of the study. Those who were interested in participating were invited to click on a Qualtrics link, which forwarded them to the letter of informed consent and brief survey to determine their eligibility. After giving consent electronically and being deemed eligible, Qualtrics randomly chose a video for them to review. After reading the case descriptions and before watching the interrogation videos or proceeding to the evaluations, participants were asked if they were already familiar with the case or had conflicts of interest (e.g., they were an expert witness on this case). If they indicated such a conflict existed, they were assigned a different video; this occurred five times. Participants then evaluated the interrogations using the assessment form described in the Measures section. Participants were then thanked, debriefed, and forwarded to a page to collect their compensation and to volunteer participant referrals for the study.

IIAI™ Measures. Two different sets of IIAI™ measures for each domain were used. The first set of measures of this study were the coercion scores produced by the full instrument. At the time of publication, I had analyzed and scored 162 interrogation videos using the IIAI™,

and observed strong skewness and kurtosis for our domains (see Table 4.1). As such strong skewness may interfere with inferential statistics using those measures, a common solution is to perform a logarithmic transformation on the data (Berry & Feldman, 1985; Field, 2009); I did so here.

The second set of measures were the tactic counts per domain. Although these were not the central measures that I was ultimately seeking to validate, the reason that the tactic counts were compared to the experts' ratings was the fact that the IIAI™ was currently in use as a guided professional judgment tool that did not reference any psychometric scoring. Therefore, validating the tactic counts as measures of coercion could be taken as a measure of convergent validity between SSEs' ratings and the IIAI™ as it was being used. Secondly, if the coercion scores did not correlate we could look at the tactic counts to see if we are even measuring the right items, or if it is a matter of measuring the right items but the scores not being representative of the coerciveness of those items.

Expert Ratings. Participants were provided a series of Likert scales on which to rate the coercive pressures of the interrogation. Response options ranged from 1 (*Not at all*) to 10 (*To a very large extent*), with a mid-point of 5 (*To a moderate extent*). The first question gave my operational definition of coercion and asked participants to evaluate the extent to which the interrogation was coercive. There were also five questions that reflected the domains of the IIAI™. Participants were asked the extent to which the investigator(s) (1) engaged in rapport building with the suspect; (2) leveraged evidence against the suspect; (3) attempted to emotionally provoke or manipulate the suspect, (4) engaged in maximization by making direct accusations, dismissing denials and objections, emphasizing the certainty of guilt or conviction; and (5) made appeals to the suspect's self-interest, offered excuses or justifications, or blamed others for the offence. Finally, participants were asked to what extent the suspect appeared

frightened, anxious, or otherwise under duress. This study was not preregistered. The data associated with this study is available through the Center for Open Science at <https://osf.io/r27uq>.

4.3 Results

The IIAITM measures were treated as the dependent variables in a series of bivariate correlational analyses with the experts' ratings. The IIAITM measures of Rapport Building for a given interrogation were paired with an expert's rating of that interrogation's level of rapport building, and so on for all of the domains.

I first correlated SSEs' total coercion ratings with those interrogations' transformed total IIAITM tactic scores. The correlation was moderate $r = .41$, $F(1, 34) = 6.735$, $p = .014$. I then conducted bivariate correlational analyses for the SSEs' ratings for each domain against IIAITM domain scores. The correlational analysis measuring the relationship between the transformed Rapport Building scores and SSEs' ratings was significant, but weak, $r = .33$, $F(1, 34) = 4.282$, $p = .046$. I found strong correlations between experts' ratings of Evidence-Based, $r = .58$, $F(1, 34) = 17.376$, $p < .001$, Social and Psychological Manipulation, $r = .66$, $F(1, 34) = 25.727$, $p < .001$, Consequence-Based, $r = .64$, $F(1, 34) = 24.128$, $p < .001$ and Minimization, $r = .48$, $F(1, 34) = 9.936$, $p = .003$, IIAITM scores. Finally, suspect duress scores correlated weakly with SSEs' ratings of suspect duress, $r = .30$, $F(1, 35) = 3.499$, $p = .013$. These results are summarized in Table 4.2.

I then conducted a series of bivariate correlational analyses comparing IIAITM tactic counts for each interrogation to experts' ratings of those interrogations. The total tactic counts showed a moderate relationship with expert's ratings of overall coercion $r = .41$, $F(1, 34) = 6.81$, $p = .013$. Correlations were strong for Evidence-Based tactic counts, $r = .51$, $F(1, 34) =$

12.233, $p = .001$, Social and Psychological Manipulation, $r = .57$, $F(1, 34) = 16.048$, $p < .001$, Consequence-Based tactic counts $r = .73$, $F(1, 34) = 38.359$, $p < .001$, and Minimization, $r = .53$, $F(1, 34) = 12.967$, $p = .001$. For Rapport and Relationship Building the relationship was nonsignificant, $r = .15$, $F(1, 34) = .733$, $p = .385$. The relationship between suspect behavior counts and SSEs' ratings of duress was nonsignificant $r = .29$, $F(1, 34) = 3.282$, $p = .079$.

4.4 Discussion

The correlation analyses of IIAI™ scores with SSEs' ratings ranged from strong to weak and nonsignificant. The instrument's measures of Rapport Building clearly differ from that of experts in the field, as did those of the suspects' level of duress. Further refinement and study are required to validate these aspects of the IIAI™. On the other hand, the IIAI™ appears to tap into the coercive use of evidence, the leveraging of consequences, and minimization as they are currently conceptualized by SSEs. The same pattern of results was found for the raw tactic counts. Thus, the IIAI™ appears to be proficient at capturing the basic components of accusatory interrogations: confront with evidence and perceptions of proof, leverage consequences, and allow a way to save face by employing minimization techniques (Inbau et al., 2013).

Chapter 5 General Discussion

Over the course of three studies, I have laid a solid empirical framework for the first standardized psychological instrument for measuring coercion in forensic interrogations. Those who serve as expert witnesses in cases of disputed confessions have indicated to me during interviews that more precise and objective measures are needed when litigating confession evidence in court, and would welcome a standardized instrument (Kaplan, manuscript in preparation). Providing experts with a firmly established collection of items to remain aware of

can ameliorate selective attention by ensuring that experts do not overlook any important factors when performing extremely complex tasks (Gawande, 2010), such as evaluating an interrogation for coercion and risk factors for false confession. By providing a checklist of items with standard definitions it may also reduce some of the cognitive biases inherent in socially-based judgments. Take, for example, the simple occurrence of an investigator accusing a suspect of lying; this item is referred to as *Information Derogation* and is clearly defined and operationalized within our instrument. From the IIAI™ user's perspective, it does not matter whether they believe that the suspect was lying or if they felt that the investigator was justified in making that accusation. The user's feelings towards the criminal offense and the imperatives of those who retained them should not interfere with the ability to recognize and code the tactic. Organizing findings based on an empirically established framework, as opposed to subjective clinical judgment, may therefore serve to reduce biases and adversarial allegiance as it has in the field of risk assessment (Gowensmith et al., 2010; Murrie & Boccaccini, 2015).

I also developed a standardized sample of 162 coded interrogations². The videotapes used to create this sample pertain to interrogations that took place in North America between 1998 and 2020, ranged from 8 to 491 minutes ($M = 73.38$, $SD = 67.31$), and included a variety of felony/indictable offenses. The file contains 116 variables and is available to interested researchers through the Center for Open Science (<https://osf.io/r27uq>). Newer versions will be periodically added as the sample continues to grow. Allowing experts to compare data from a given interrogation to a much larger standardized sample of interrogations may help reduce the anchoring effect and assist experts in making comparisons about the coerciveness of an interrogation to norms. Thus, distilling the past 40 years of research into a systemized instrument for the first time allows expert witnesses to provide quantifications of coercive

pressure based on the aggregate knowledge of the social science on interrogations and confessions.

In addition to guiding expert witnesses' judgments, the IIAI™ will be useful in future research for coding interrogations by providing common sets of metrics and measures by which to compare results. Users, however, should be mindful of the limitation that inter-rater reliability varied among the domains. The IIAI™ produces measures for 90 different variables relating to the interrogation coded, including an exact record of interrogation tactics used and suspect behaviors observed. These measures are spread out across numerous domains, allowing for the quantification of the amount of coercive pressure being applied and how it was being applied. The various measures of the IIAI™ may be analyzed by themselves, in relation to one another, or in relation to other considerations such as the location, the suspects' demographics (e.g., gender, race), the crime(s) under investigation, etc.

5.1 Limitations

Although during pilot testing I established good-to-excellent levels of interrater reliability for all of our domains, in Study 2 there was considerable disagreement for Rapport Building, Social and Psychological Manipulation, and Suspect Behaviors. Some individual items such as Evidence Bait, Past Acts, and the Alternate Question also showed low interrater reliability in Study 2. Part of the reason for this discrepancy between the pilot study and Study 2 may be the fact that the research assistants from the pilot study had much more experience using the instrument by the end of the five sets than did the research assistants recruited for Study 2. Therefore, the practice coding of at least five interrogation videos will be added to the training, and the interrater reliability of the IIAI™ will be better established through future study. Regardless of improvements made, completely perfect interrater reliability is unlikely as all

psychological instruments show at least some variation between users (Blais et al., 2017; Edens et al., 2016). The instrument is still vulnerable to human error, and users may not always be attentive to all tactics. Each trained IIAI™ user may have their own idiosyncrasies and the same interrogation evaluated by two experts will invariably show some differences.

Rapport Building was a weak spot in both Studies 2 and 3. Unlike concepts such as minimization and maximization found in accusatory interrogations, the measurement of rapport building has been vague and elusive in the social sciences (Gabbert et al., 2021). In my studies, the construct of rapport building was not so easily captured in a set of discretely coded items; this method of measurement does not take into account non-verbal techniques or concepts such as *mutual attentiveness* (showing caring interest), *positivity* (positive affect), and *coordination* (shared understanding; Tickle-Degnen & Rosenthal, 1990) that experts may have been attentive to in making their ratings. In practise, this means the IIAI™ may not be fully suited to capturing social pressures that may be present in rapport-based methods of questioning, such as PEACE (Centrex, 2004). Making refinements to our measures of Rapport Building will be of focus for future revisions of the IIAI™.

There are also several important expert witness functions that are not aided by the IIAI™. It is not a method of evaluating an investigator's proficiency at conducting investigative interviews. The IIAI™ is solely for evaluating coercion in interviews and interrogations, and does not provide any framework for evaluating the larger investigation or police tunnel vision. It may assist in confession contamination analysis by marking on its timeline when investigators raise evidence, but it is not a tool specifically meant for evaluating the reliability of the information gained from interviews and interrogations.

Study 3 employed a relatively small number of participants who may not fully reflect the state of the social science knowledge on coercion and interrogations, particularly considering that nearly half had never served as expert witnesses. Given the small sample size we also opted not to perform adjustments of our alpha levels; we were interested in the strength of the relationships and believed it would be overly stringent to apply such a correction given the relatively small sample size.

The IIAITM was designed to capture coercive pressure in interrogations and thereby serve as a measure of the risk of coerced-compliant false confessions. In Study 3 I established a basic level of convergent validity with how experts currently working in the criminal justice system conceptualize coercive pressure, but I have not established criterion validity with the outcome of a suspect making a false confession. With a sample of videotaped interrogations resulting in known false confessions an ROC analysis could potentially be performed to demonstrate that those interrogations with higher scores are more likely to result in false confessions, and this represents an area of future study. Not having establishing criterion validity does not represent a fatal flaw since an expert may not directly opine as to the ultimate question of a confession's veracity or the guilt of a defendant (Marion et al., 2019). Further, the majority of the items included in the instrument such as explicit threats, promises, minimization and maximization techniques, etc., have been independently confirmed to be risk factors for false confession through experimental research, archival research, and meta-analyses (Leo & Drizin, 2004; Leo & Ofshe, 1997; Meissner et al; 2012; 2014).

In Study 1 I removed eight tactics due to experts rating them as not having coercive properties. While certain tactics such as depersonalizing the situation may not be coercive when used in isolation, when paired with other tactics (such as a direct accusation) they may have a

coercive impact. In such cases where a coercive tactic is augmented by some other occurrence not directly captured by the IIAI™ it may be mentioned in the evaluation, but it would not impact the instrument's scoring.

The IIAI™ tactic scores as derived from the results of Study 1 and used in Study 3 are not perfect measures. Surveying experts was a proficient method of validating the items that had coercive properties and discarding those that did not, but it may not have been an appropriate method to determine the coercive magnitude of each item. Ultimately, science is not a democracy where a poll is taken and an averaged middle ground of opinions is accepted as truth, and the entire basis of the scoring matrix may be predicated on shaky ground. The SSEs' ratings from Study 3 are also likely less than perfect reflections of coercive pressure. The experts' ratings may have been affected by all of the biases and shortcomings of unguided judgments discussed above. The anchoring effect in particular may have played a large role in that the IIAI™ measures were produced by a standardized instrument, whereas the expert ratings were not only produced by unguided judgment, but the unguided judgment of 36 different individuals each theoretically with their own anchoring points. If each of the 36 videos had been reviewed and judged relatively to one another by a single expert, then the correlations may have been different.

Throughout the thesis to this point, I have referred to the IIAI™ as an instrument for measuring coercion and risk factors for coerced compliant false confession, yet Study 1 and 3 only asked participants to rate the former. I maintain that the IIAI™ will still assist the courts in adjudicating disputed confessions decisions. In order for a confession to constitute admissible inculpatory evidence it must be made voluntarily (*R v. Piche*, 1971; Smith et al., 2012). By definition, a coerced statement is not voluntary and, therefore, knowing that the items had been

rated as coercive by experts is useful knowledge for the trier of law in determining admissibility. Knowledge about the amount of coercion that elicited a confession is also useful information to jurors in weighing veracity; the link between coercive interrogation practises and false confession is so strong and direct (see Drizin & Leo, 2004; Kassin et al., 2010; Scherr et al., 2020) that evaluating the amount of psychological coercion behind a statement is one of the main criteria for evaluating the weight of confession evidence (Kassin, 1997).

The IIAI™ takes into account suspect risk factors, such as mental illness and intellectual disability, but it is not a psychometric instrument for assessing a suspect's mental capacity. Such diagnoses must be made by psychiatrists or clinical psychologists; their findings may then be incorporated into the IIAI™ evaluation. Determinations of personal risk factors may be informed by materials provided by retaining attorneys such as clinician's reports, MMPI or Wechsler Intelligence Scale measures, past histories of psychiatric hospitalization, etc., but making the determination as to whether those personal risk factors are present is not at the IIAI™ user's discretion based solely on what they observed in the videotaped interrogation.

The IIAI™ was developed in North America by Canadians and Americans using a sample of interrogation videos which all came from North American law enforcement agencies. The basic principles and the items reflecting those principles should be mostly universal because accusatory approaches are used by a variety of law enforcement agencies around the globe. In addition to being widely used in North America, accusatory and hybrid methods of interrogations, which are premised on choice architecture, are taught to and employed by police internationally including in Japan (Takagi, 2012; Wachi & Lamb, 2019), Korea (Reid, 2015), South Africa (Starr, 2019), and across the Middle East (Inbau et al., 2013) and Latin America (Wicklander-Zulawski, 2020). The instrument has, however, yet to be tested on interrogations

from outside of North American jurisdictions and should not be used to do so until its reliability and validity can be established using international samples.

5.2 Conclusion

Over the course of three studies, I have laid the empirical foundations for the IIAI™. In Study 1 I was able to greatly reduce a large number of items down to a workable set of items pertinent to the assessment of coercion in police interrogations that were internally reliable. After making significant edits and refinements during pilot testing, at the end of that process I had developed a usable instrument, and then tested its interrater reliability in Study 2. Finally, in Study 3 I established a basic level of convergent validity between most of the instrument's measures and the evaluations of coercive pressure of the forensic social science community. Further refinements and testing are needed in some areas, as discussed in the Limitations section, but the basic framework that I developed has largely stood up to empirical scrutiny.

The field of forensic psychology has been moving towards greater standardization for decades (Andrews et al., 2006), and my instrument is a natural step in this process. Equipping forensic social scientists and expert witnesses with a measurement of coercion based on psychological standards and relevant to criminal and procedural law will advance both the state of forensic social science and the effectiveness of the criminal justice system. The instrument is meant to standardize how experts evaluate interrogations in the criminal justice system and provide more comprehensive and objective analyses of confession evidence.

References

- Alceste, F., J Luke, T., D Redlich, A., Hellgren, J., D Amrom, A., & M Kassin, S. (2021). The psychology of confessions: A comparison of expert and lay opinions. *Applied Cognitive Psychology, 35*, 39-51. *Psychology, 35*, 39-51. <https://doi.org/10.1002/acp.3735>
- Andrews, D. A., Bonta, J., & Wormith, J. S. (2006). The recent past and near future of risk and/or need assessment. *Crime & delinquency, 52*, 7-27. <https://doi.org/10.1177/0011128705281756>
- Bang, B. L., Stanton, D., Hemmens, C., & Stohr, M. K. (2018). Police recording of custodial interrogations: A state-by-state legal inquiry. *International Journal of Police Science & Management, 20*, 3-18. <https://doi.org/10.1177/1461355717750172>
- Baumeister, R. F., Vohs, K. D., & Tice, D. M. (2007). The strength model of self-control. *Current directions in psychological science, 16*, 351-355. <https://doi.org/10.1111/j.1467-8721.2007.00534.x>
- Berry, W. D., & Feldman, S. (1985). The multiple regression model: a review. *Multiple regression in practice. Newbury Park, CA: Sage University Paper*, 9-17.
- Blagrove, M. (1996). Effects of length of sleep deprivation on interrogative suggestibility. *Journal of Experimental Psychology: Applied, 2*, 48-59. <http://dx.doi.org/10.1037/1076-898X.2.1.48>
- Blair, J. (2005). A test of the unusual false confession perspective using cases of proven false confessions. *Criminal Law Bulletin, 41*, 127-144.
- Blais, J., Forth, A. E., & Hare, R. D. (2017). Examining the interrater reliability of the Hare Psychopathy Checklist—Revised across a large sample of trained raters. *Psychological Assessment, 29*, 762-775. <https://doi.org/10.1037/pas0000455>

- Borchard, E. M. (1932). *Convicting the Innocent*. Garden City, NY: Garden City Publishing Company.
- Bram v. United States*, 168 US 532 (1897).
- Bull, R., & Soukara, S. (2010). Four studies of what really happens in police interviews. In G. D. Lassiter & C. A. Meissner (Eds.), *Police interrogations and false confessions: Current research, practice, and policy recommendations* (pp. 81–95). Washington, DC: American Psychological Association.
- Canadian Charter of Rights and Freedoms*, s 7-12, Part I of the *Constitution Act* (1982), being Schedule B to the *Canada Act 1982*.
- Carr, D. (2015). *The RCMP Phased Interview Model for Suspects: Generating information through dialogue*. Ottawa, ON: Royal Canadian Mounted Police.
- Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*, 6, 284–290. <https://doi.org/10.1037/1040-3590.6.4.284>
- Cleary, H. M. D. (2017). Applying the lessons of developmental psychology to the study of juvenile interrogations: New directions for research, policy, and practice. *Psychology, Public Policy, and Law*, 23(1), 118–130. Retrieved from <http://dx.doi.org/10.1037/law0000120>
- Costanzo, M., & Leo, R. A. (2007). Research and expert testimony on interrogations and confessions. In M Costanzo (Ed.) *Expert psychological testimony for the courts*, (pp. 69-98) Hove, UK: Psychology Press.
- Crough, Q., Dion Larivière, C., Snow, M. D., & Eastwood, J. (2022). Reflections on the nature of rapport within suspect interviews. *Current Issues in Criminal Justice*, 34, 219-228.

- Davis, D., & Leo, R. A. (2012). Interrogation-related regulatory decline: Ego depletion, failures of self-regulation, and the decision to confess. *Psychology, Public Policy, and Law*, *18*, 673-704. <https://doi.org/10.1037/a0027367>
- Davis, D., & Leo, R. A. (2013). Acute suggestibility in police interrogation: Self-regulation failure as a primary mechanism of vulnerability. In A. M. Ridley, F. Gabbert, and D. J. La Rooy (Eds.), *Suggestibility in legal contexts: Psychological research and forensic implications* (pp. 171–195). Oxford, UK: Wiley-Blackwell.
- DePaulo, B. M., Lindsay, J. J., Malone, B. E., Muhlenbruck, L., Charlton, K., & Cooper, H. (2003). Cues to deception. *Psychological bulletin*, *129*, 74–118. <https://doi.org/10.1037/0033-2909.129.1.74>
- Drizin, S., & Leo, R. (2004). The problem of false confessions in a post-DNA world. *North Carolina Law Review*, *82*, 891-1007.
- Edens, J. F., Penson, B. N., Ruchensky, J. R., Cox, J., & Smith, S. T. (2016). Interrater reliability of Violence Risk Appraisal Guide scores provided in Canadian criminal proceedings. *Psychological Assessment*, *28*, 1543–1549. <https://doi.org/10.1037/pas0000278>
- Evans, D. R., Eisenlohr-Moul, T. A., Button, D. F., Baer, R. A., & Segerstrom, S. C. (2014). Self-regulatory deficits associated with unpracticed mindfulness strategies for coping with acute pain. *Journal of Applied Social Psychology*, *44*, 23-30.
- Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. <https://doi.org/10.3758/BF03193146>
- Field, A. (2009). *Discovering statistics using SPSS*. Thousand Oaks, CA: Sage Publications.

- Frenda, S. J., Berkowitz, S. R., Loftus, E. F., & Fenn, K. M. (2016). Sleep deprivation and false confessions. *Proceedings of the National Academy of Sciences*, *113*(8), 2047–2050.
<https://doi.org/10.1073/pnas.1521518113>
- Gabbert, F., Hope, L., Luther, K., Wright, G., Ng, M., & Oxburgh, G. (2021). Exploring the use of rapport in professional information-gathering contexts by systematically mapping the evidence base. *Applied Cognitive Psychology*, *35*, 329-341.
<https://doi.org/10.1002/acp.3762>
- Gawande, A. (2010). *The Checklist manifesto*. Penguin Books India.
- Gowensmith, W. N., Murrie, D. C., & Boccaccini, M. T. (2010). Field reliability of competence to stand trial opinions: How often do evaluators agree, and what do judges decide when evaluators disagree? *Law and Human Behavior*, 1-13. <https://doi.org/10.1037/h0093958>
- Grove, W. M., Zald, D. H., Lebow, B. S., Snitz, B. E., & Nelson, C. (2000). Clinical versus mechanical prediction: a meta-analysis. *Psychological Assessment*, *12*, 19-30.
<https://doi.org/10.1037/1040-3590.12.1.19>
- Gudjonsson, G. H. (2003). *The psychology of interrogations and confessions*. Chichester, UK: Wiley.
- Gudjonsson, G. H. (2010). Psychological vulnerabilities during police interviews. Why are they important? *Legal and Criminological Psychology*, *15*, 161–175.
<https://doi.org/10.1348/135532510X500064>
- Gudjonsson, G. H. (2021). The science-based pathways to understanding false confessions and wrongful convictions. *Frontiers in Psychology*, *12*, 308.
<https://doi.org/10.3389/fpsyg.2021.633936>

- Hanson, R. K., & Morton-Bourgon, K. E. (2009). The accuracy of recidivism risk assessments for sexual offenders: a meta-analysis of 118 prediction studies. *Psychological assessment, 21*, 1-21. <https://doi.org/10.1037/a0014421>
- Henkel, L. A., Coffman, K. A., & Dailey, BA, E. M. (2008). A survey of people's attitudes and beliefs about false confessions. *Behavioral Sciences & the Law, 26*, 555-584. <https://doi.org/10.1002/bsl.826>
- Hopt v. Utah*, 110 US 574 (1884).
- Inbau, F., Reid, J., Buckley, J., & Jayne, B. (2013). *Criminal interrogation and confessions*. 5th ed. Burlington, MA: Jones & Bartlett Learning.
- Jones, A. M., & Penrod, S. (2016). Can expert testimony sensitize jurors to coercive interrogation tactics? *Journal of Forensic Psychology Practice, 16*, 393-409. <https://doi.org/10.1080/15228932.2016.1232029>
- Kaplan, J. (in preparation). Expert testimony on disputed confessions: Bridging theory and practise.
- Kaplan, J. & Cutler, B. L. (2021). Co-occurrences among interrogation tactics in actual police investigations. *Psychology, Crime, & Law, 1-19*. <https://doi.org/10.1080/1068316X.2021.1876051>
- Kaplan, J., Cutler, B.L., Leach, A.M., Eastwood, J., & Marion, S. (2019). Evaluating coercion in suspect interviews and interrogations. In B. Bornstein & M. Miller's (Eds.) *Advances in Psychology and Law* (pp. 1-40). New York, NY: Springer.
- Kaplan, J., Cutler, B. L., Leach, A.M., Marion, S., & Eastwood, J. (2020). Perceptions of coercion: Comparing laypersons and experts. *Psychology, Crime, and Law, 384-401*. <https://doi.org/10.1080/1068316X.2019.1669597>

- Kaplan, J., Cutler, B. L., Leach, A.M., Marion, S., & Eastwood, J. (2020). *Development of the Interview and Interrogation Assessment Instrument study one survey data* [Data set]. Center for Open Science.
https://osf.io/n6hgw?view_only=028d8d1a9c14492095721e8623ce7b75
- Kaplan, J., Cutler, B. L., Leach, A.M., & Eastwood, J. (2022). *Development of the Interview and Interrogation Assessment Instrument study three validation study data* [Data set]. Center for Open Science.
https://osf.io/kcqp2?view_only=028d8d1a9c14492095721e8623ce7b75
- Kaplan, J., Rumschik, D., Woolridge, L. R., Snow, M. D., Rodriguez Steen, L. (2022). *Interview and Interrogation Assessment Instrument standardized sample* [Data set]. Center for Open Science. https://osf.io/kchfq/?view_only=028d8d1a9c14492095721e8623ce7b75
- Kassin, S. M. (1997). The psychology of confession evidence. *American Psychologist*, *52*, 221-233. <https://doi.org/10.1037/0003-066X.52.3.221>.
- Kassin, S. M., Drizin, S. A., Grisso, T., Gudjonsson, G. H., Leo, R. A., & Redlich, A. D. (2010). Police-induced confessions: Risk factors and recommendations. *Law and Human Behavior*, *34*, 3-38. <https://doi.org/10.1007/s10979-009-9188-6>
- Kassin, S. M., Leo, R. A., Meissner, C. A., Richman, K. D., Colwell, L. H., Leach, A. M., & La Fon, D. (2007). Police interviewing and interrogation: A self-report survey of police practices and beliefs. *Law and Human Behavior*, *31*, 381-400.
<https://doi.org/10.1007/s10979-006-9073-5>
- Kassin, S. M., & McNall, K. (1991). Police interrogations and confessions: Communicating promises and threats by pragmatic implication. *Law and Human Behavior*, *15*, 233–251.
<https://doi.org/10.1007/BF01061711>

- Kassin, S. M., Redlich, A. D., Alceste, F., & Luke, T. J. (2018). On the general acceptance of confessions research: Opinions of the scientific community. *American Psychologist, 73*, 63-80. <https://doi.org/10.1037/amp0000141>
- Kelly, C. E., Miller, J. C., & Redlich, A. D. (2016). The dynamic nature of interrogation. *Law and Human Behavior, 40*, 295-309. <https://doi.org/10.1037/lhb0000172>
- Kelly, C. E., Miller, J. C., Redlich, A. D., & Kleinman, S. M. (2013). A taxonomy of interrogation methods. *Psychology, Public Policy, and Law, 19*, 165–178. <https://doi.org/10.1037/a0030310>
- Kelly, C. E., Redlich, A. D., & Miller, J. C. (2015). Examining the meso-level domains of the interrogation taxonomy. *Psychology, Public Policy, and Law, 21*, 179–191. <https://doi.org/10.1037/law0000034>
- King, L., & Snook, B. (2009). Peering inside a Canadian interrogation room: An examination of the Reid Model of interrogation, influence tactics, and coercive strategies. *Criminal Justice and Behavior, 36*, 674–694. <https://doi.org/10.1177/0093854809335142>
- King v. Warickshall*, 168 Eng. Rep. 234, 235 (K.B. 1783).
- Klaver, J. R., Lee, Z., & Rose, V. G. (2008). Effects of personality, interrogation techniques and plausibility in an experimental false confession paradigm. *Legal and Criminological Psychology, 13*, 71-88. <https://doi.org/10.1348/135532507X193051>
- Knowles, E. S., & Linn, J. A. (Eds.). (2004). *Resistance and persuasion*. Psychology Press.
- Knowles, E. S., & Riner, D. D. (2007). Omega approaches to persuasion: Overcoming resistance. In Pratkanis, A. (Eds.) *The science of social influence: Advances and future progress*, (pp. 83-114). New York, NY, US: Psychology Press.

- Krauss, D., & Lieberman, J. (2007). Expert testimony on risk and future dangerousness. *Expert Testimony for the Courts*, 227-250.
- Krauss, D. A., & Sales, B. D. (2001). The effects of clinical and scientific expert testimony on juror decision making in capital sentencing. *Psychology, Public Policy, and Law*, 7, 267-310. <https://doi.org/10.1037/1076-8971.7.2.267>
- Leo, R. A. (1996). Inside the interrogation room. *The Journal of Criminal Law and Criminology*, 86, 266-303.
- Leo, R. A. (2008). *Police interrogation and American justice*. Harvard University Press.
- Leo, R. A., & Liu, B. (2009). What do potential jurors know about police interrogation techniques and false confessions? *Behavioral Sciences & the Law*, 27, 381-399. <https://doi.org/10.1002/bsl.872>
- Leo, R. A., & Ofshe, R. J. (1998). The consequences of false confessions: Deprivations of liberty and miscarriages of justice in the age of psychological interrogation. *The Journal of Criminal Law and Criminology*, 88, 429-496.
- Lidz, C. W., Mulvey, E. P., Hoge, S. K., Kirsch, B. L., Monahan, J., Bennett, N. S., ... & Roth, L. H. (1997). The validity of mental patients' accounts of coercion-related behaviors in the hospital admission process. *Law and Human Behavior*, 21, 361-376. <https://doi.org/10.1023/a:1024803102017>
- Luna, S. (2022). Defining coercion: An application in interrogation and plea negotiation contexts. *Psychology, Public Policy, and Law*. <https://doi.org/10.1037/law0000345>
- Malinski v. New York*, 324 US 401, 404, (1945).

- Malloy, L. C., Shulman, E. P., & Cauffman, E. (2014). Interrogations, confessions, and guilty pleas among serious adolescent offenders. *Law and Human Behavior, 38*(2), 181–193. Retrieved from <https://doi.org/10.1037/lhb0000065>
- Marion, S., Kaplan, J., & Cutler, B.L. (2019). Expert testimony. In N. Brewer & A. Douglass (Eds.), *Psychological Science and the Law*. New York, NY: Guilford Press.
- McAuliff, B. D., & Arter, J. L. (2016). Adversarial allegiance: The devil is in the evidence details, not just on the witness stand. *Law and Human Behavior, 40*, 524-535. <https://doi.org/10.1037/lhb0000198>
- Meissner, C. A., Redlich, A. D., Bhatt, S., & Brandon, S. (2012). Interview and interrogation methods and their effects on true and false confessions. *Campbell Systematic Reviews, 8*, 4-52. 4-52. <https://doi.org/10.4073/csr.2012.13>
- Meissner, C. A., Redlich, A. D., Michael, S. W., Evans, J. R., Camilletti, C. R., Bhatt, S., & Brandon, S. (2014). Accusatorial and information-gathering interrogation methods and their effects on true and false confessions: A meta-analytic review. *Journal of Experimental Criminology, 10*, 459-486. <https://doi.org/10.1007/s11292-014-9207-6>
- Melzer, N. (2018). Seventieth anniversary of the Universal Declaration of Human Rights: reaffirming and strengthening the prohibition of torture and other cruel, inhuman or degrading treatment or punishment. Interim report of the Special Rapporteur on torture and other cruel, inhuman or degrading treatment of punishment.
- Miranda v. Arizona*, 384 U.S. 436. (U.S.S.C., 1966).
- Monahan, J. (1981). *Predicting violent behavior: An assessment of clinical techniques*. Beverly Hills, CA: Sage Publications.

- Munsterberg, H. (1908). *On the witness stand: Essays on psychology and crime*. New York, NY: Doubleday.
- Muraven, M., & Baumeister, R. F. (2000). Self-regulation and depletion of limited resources: Does self-control resemble a muscle? *Psychological Bulletin*, *126*, 247-259.
<https://doi.org/10.1037/0033-2909.126.2.247>
- Murrie, D. C., & Boccaccini, M. T. (2015). Adversarial allegiance among expert witnesses. *Annual Review of Law and Social Science*, *11*, 37-55. <http://dx.doi.org/10.1146/annurev-lawsocsci-120814-121714>
- Murrie, D. C., Boccaccini, M. T., Turner, D. B., Meeks, M., Woods, C., & Tussey, C. (2009). Rater (dis)agreement on risk assessment measures in sexually violent predator proceedings: Evidence of adversarial allegiance in forensic evaluation? *Psychology, Public Policy, and Law*, *15*, 19-53. <https://doi.org/10.1037/a0014897>
- Murrie, D. C., Boccaccini, M. T., Guarnera, L. A., & Rufino, K. A. (2013). Are forensic experts biased by the side that retained them?. *Psychological Science*, *24*, 1889-1897.
<https://doi.org/10.1177/0956797613481812>
- Neal, T., & Grisso, T. (2014). The cognitive underpinnings of bias in forensic mental health evaluations. *Psychology, Public Policy, and Law*, *20*, 200-211.
<https://doi.org/10.1037/a0035824>
- Nozick, R. (1969). Coercion. In S. Morgenbesser, P. Suppes, and M. White (Eds.), *Philosophy, science, and method: Essays in honor of Ernest Nagel* (pp. 440–472). St. Martin's Press.
- Ofshe, R. J., & Leo, R. A. (1997). The social psychology of police interrogation: The theory and classification of true and false confessions. *Studies in Law Politics and Society*, *16*, 189-254.

- Pearse, J., & Gudjonsson, G. H. (1999). Measuring influential police interviewing tactics: A factor analytic approach. *Legal and Criminological Psychology, 4*, 221-238.
<https://doi.org/10.1348/135532599167860>
- Pepson, M. D., & Sharifi, J. N. (2010). Lego v. Twomey: The improbable relationship between an obscure Supreme Court decision and wrongful convictions. *American Criminal Law Review, 47*, 1185-1250.
- Philips, M. (1984). Are coerced agreements involuntary? *Law and Philosophy, 3*, 133-145.
- R v. Oickle*, 2 SCR 3, SCR 38. (S.C.C., 2000).
- R v. Piche*, SCR 23, (S.C.C., 1971).
- Redlich, A. D. (2004). Law & psychiatry: Mental illness, police interrogations, and the potential for false confession. *Psychiatric Services, 55*, 19–21.
<https://doi.org/10.1176/appi.ps.55.1.19>
- Redlich, A. D. (2007). Double jeopardy in the interrogation room: Young age and mental illness. *American Psychologist, 62*, 609–611. <https://doi.org/10.1037/0003-066X62.6.609>
- Redlich, A. D. (2009). The susceptibility of juveniles to false confessions and false guilty pleas. *Rutgers Law Review, 62*, 943–957.
- Redlich, A. D., Kulish, R., & Steadman, H. J. (2011). Comparing true and false confessions among persons with serious mental illness. *Psychology, Public Policy, and Law, 17*, 394–419. <http://dx.doi.org/10.1037/a0022918>
- Reid Institute. (2015). International research validates the core elements of the Reid Technique.
Retrieved from: <http://www.reid.com/pdfs/20150511.pdf>
- Rufino, K. A., Boccaccini, M. T., Hawes, S. W., & Murrie, D. C. (2012). When experts disagreed, who was correct? A comparison of PCL-R scores from independent raters and

- opposing forensic experts. *Law and Human Behavior*, 36, 527-537.
<https://doi.org/10.1037/h0093988>
- Schatz, S. J. (2018). Interrogated with intellectual disabilities: The risks of false confession. *Stanford Law Review*, 70, 643-690.
- Scherr, K. C., Redlich, A. D., & Kassin, S. M. (2020). Cumulative disadvantage: A psychological framework for understanding how innocence can lead to confession, wrongful conviction, and beyond. *Perspectives on Psychological Science*, 15, 353-383.
<https://doi.org/10.1177/1745691619896608>
- Schulhofer, S. J. (2001). Miranda, Dickerson, and the puzzling persistence of Fifth Amendment exceptionalism. *Michigan Law Review*, 99, 941-957.
- Smith, S., Stinson, V., & Patry, M. (2012). Confession evidence in Canada: psychological issues and legal landscapes. *Psychology, Crime & Law*, 18, 317-333.
<https://doi.org/10.1080/1068316X.2010.486380>
- Starr, D. (June 13, 2019). This psychologist explains why people confess to crimes they didn't commit. *Science Magazine*. Retrieved from:
<https://www.sciencemag.org/news/2019/06/psychologist-explains-why-people-confess-crimes-they-didn-t-commit>
- Soree, N. (2005). When the innocent speak: False confessions, constitutional safeguards, and the role of expert testimony. *American Journal of Criminal Law*, 32, 191-263.
- Takagi, K. (2012). Toward a scientific approach to investigative interviewing: Lessons from a false confession in the Ashikaga case. *Japanese Journal of Law and Psychology*, 12, 16-22. <https://doi.org/10.5926/arepj.52.115>

- Thomas III, G. C., & Leo, R. A. (2009). *Confessions of guilt: From torture to Miranda and beyond*. Oxford, UK: Oxford University Press.
- Thomas III, G. C., & Leo, R. A. (2012). Law of Police Interrogation. Encyclopedia of Criminology & Criminal Justice (Springer 2013), Univ. of San Francisco Law Research Paper.
- Tickle-Degnen, L., & Rosenthal, R. (1990). The nature of rapport and its nonverbal correlates. *Psychological Inquiry, 1*, 285-293. https://doi.org/10.1207/s15327965pli0104_1
- Trainum, J. L. (2016). *How the police generate false confessions: An inside look at the interrogation room*. Lanham, MD: Rowman & Littlefield.
- Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology, 5*, 207-232. [https://doi.org/10.1016/0010-0285\(73\)90033-9](https://doi.org/10.1016/0010-0285(73)90033-9)
- Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185*, 1124-1131. <https://doi.org/10.1126/science.185.4157.1124>
- Tversky, A., & Kahneman, D. (1986). The framing of decisions and the evaluation of prospects. In *Studies in Logic and the Foundations of Mathematics* (pp. 503-520). Elsevier.
- U.S. Constitution, Amendment II*. (1791).
- Vallano, J. P., & Schreiber Compo, N. (2015). Rapport-building with cooperative witnesses and criminal suspects: A theoretical and empirical review. *Psychology, Public Policy, and Law, 21*, 85-99. <https://doi.org/10.1037/law0000035>
- Vrij, A., Fisher, R., Mann, S., & Leal, S. (2006). Detecting deception by manipulating cognitive load. *Trends in Cognitive Sciences, 10*, 141-142. <https://doi.org/10.1016/j.tics.2006.02.003>

- Wachi, T., & Lamb, M. E. (2019). Japanese public opinion about suspect interviewing techniques. *Legal and Criminological Psychology, 24*, 55-70.
<https://doi.org/10.1111/lcrp.12139>
- Watts, S. E., & Weems, C. F. (2006). Associations among selective attention, memory bias, cognitive errors and symptoms of anxiety in youth. *Journal of Abnormal Child Psychology, 34*, 838-849. <https://doi.org/10.1007/s10802-006-9066-3>
- Wickersham Commission Report. (1931). *Report on lawlessness in law enforcement*. National Commission on Law Observance and Enforcement. Washington, DC: U.S. Government
- Wicklander-Zulawski & Associates. (2020). *Interviewing techniques*. Aurora, IL: Wicklander-Zulawski & Associates.

Footnotes

¹The HAITM was originally named and is discussed as the *Coercion Assessment Instrument* in Kaplan et al. (2019). In that chapter we introduced the early conceptualization of the instrument and attempts to operationalize coercion. The studies discussed in the current article and the analyses of the instrument's reliability and validity only appear in the current article.

²Interrogation data that came from cases in which the authors were retained as expert witnesses has been removed from the publicly available file.

Table 3.1

Pilot study interrater reliability by interrogation

	Total	Rapport	Evidence	S&P	Consequence	Min	Suspect
Set 1	.556, .790	.108, .266	.563, .794	.183, .401	.623, .832	.326, .592	
Set 2	.594, .815	.522, .766	.807, .926	.626, .834	.320, .586	.396, .663	
Set 3	.764, .907	.308, .572	.787, .917	.921, .972	.393, .660	.631, .837	.498, .665
Set 4	.870, .952	.500, .750	.919, .971	.859, .948	.773, .911	.419, .684	.581, .735
Set 5	.875, .933	.650, .788	.899, .947	.915, .956	.864, .927	.853, .920	.800, .889

Single and averaged ICC 2, k measures

Note: Coding of suspect behaviors began in set 3 of pilot testing; tactic total calculated independent of suspect data. S&P is the Social and Psychological Manipulation domain.

Table 3.2

Pilot study interrater reliability by time block

	Total	Rapport	Evidence	S&P	Consequence	Min	Suspect
Set 1	.384, .651	.147, .341	.443, .705	.183, .401	.213, .448	.408, .674	
Set 2	.329, .596	.295, .556	.397, .664	.626, .834	.378, .645	.297, .558	
Set 3	.592, .813	.448, .708	.624, .833	.921, .972	.692, .871	.388, .655	.542, .703
Set 4	.676, .862	.476, .732	.747, .899	.859, .948	.656, .851	.591, .812	.469, .663
Set 5	.741, .851	.685, .813	.779, .876	.915, .956	.713, .832	.723, .839	.618, .764

Single and averaged ICC 2, k measures

Note: Coding of suspect behaviors began in set 3 of pilot testing; tactic total calculated independent of suspect data. S&P is the Social and Psychological Manipulation domain.

Table 3.3

Study 2 interrater reliability by video and time block

	Total	Rapport	Evidence	S&P	Consequence	Min	Suspect
Video	.944, .985	.285, .615	.970, .992	.673, .892	.704, .905	.647, .880	.487, .851
Block	.856, .960	.457, .771	.925, .980	.490, .794	.573, .843	.613, .864	.316, .735

Single and averaged ICC 2, k measures

Table 3.4
*Average interrogation Environment and Suspect
 Characteristic interrater reliability all sets*

Set	Environment and Context	Suspect Risk Factors
	κ	κ
Pilot Set 1	.661	.704
Pilot Set 2	.825	.927
Pilot Set 3	.671	.769
Pilot Set 4	.568	1.000
Pilot Set 5	.947	1.000
Study 2	.719	.751

Table 4.1
IIAITM scores descriptive statistics

Measure	Pre-transformation						Post-transformation					
	<i>M</i>	<i>SD</i>	<i>Skew</i>	<i>SE Skew</i>	<i>Kurtosis</i>	<i>SE Kurtosis</i>	<i>M</i>	<i>SD</i>	<i>Skew</i>	<i>SE Skew</i>	<i>Kurtosis</i>	<i>SE Kurtosis</i>
Rapport Building	3.85	4.51	1.98	.191	4.40	.379	1.22	.84	.21	.191	-.82	.379
Evidence-Based	18.97	23.78	2.27	.191	6.05	.379	2.36	1.18	-.08	.191	-.78	.379
S&P Manipulation	9.32	19.27	3.58	.191	14.56	.379	1.31	1.37	.65	.191	-.69	.379
Consequence	39.32	57.18	3.29	.191	14.46	.379	2.85	1.46	-.39	.191	-.79	.379
Minimization	13.88	20.56	3.19	.191	12.59	.379	1.99	1.23	.01	.191	-.78	.379
Suspect Behaviors	35.41	41.76	3.45	.191	16.31	.379	3.94	1.14	-.49	.191	.45	.379
Total	79.95	108.18	3.03	.191	12.63	.379	4.53	1.56	-.93	.191	1.46	.379

N = 162

Table 4.2
Study 3 tactic counts and IIAITM scores correlated with experts' ratings

Measure	Tactic counts		IIAITM scores	
	<i>r</i>	<i>p</i>	<i>r</i>	<i>p</i>
Rapport Building	.15	.385	.33	.046
Evidence-Based	.51	.001	.58	<.001
S&P Manipulation	.57	<.001	.66	<.001
Consequences	.73	<.001	.64	<.001
Minimization	.53	.001	.48	.003
Suspect Duress	.29	.079	.30	.035
Total	.41	.013	.41	.014

N = 36

Figure 1.1
Active coding of interrogation tactics using the IIAI™

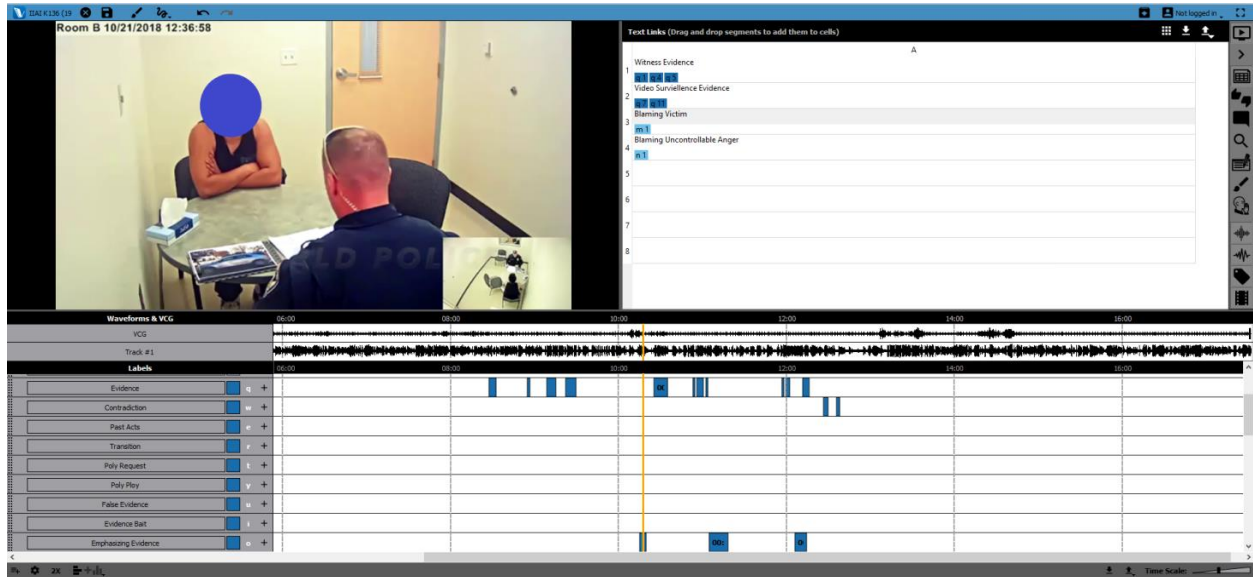


Figure 1.2
Active coding of suspect behaviors using the IIAI™

