

Deception detection: State of the art and future prospects

Jaume Masip
Universidad de Salamanca

Abstract

Background: Deception detection has been a longstanding concern throughout human history. It has also interested scientists, who have explored psychological and behavioral differences between liars and truth tellers, as well as ways to improve detection accuracy. **Method:** In recent years, substantial advances have been made in the field. Some of these advances are briefly reviewed in the current article. **Results:** A description is provided of (a) research and contemporary theories on how people (try to) detect deception; (b) recent advances on strategic interviewing to detect deception; (c) the integrative findings of recent meta-analyses on systematic verbal lie detection approaches; and (d) several important aspects concerning psychophysiological detection of deception. Also, some emerging trends and research needs for the future are outlined at the end of the article. **Conclusions:** Deception detection research is a lively and dynamic area of applied psychology that has experienced substantial developments in recent times. Much (though not all) of these research efforts have focused on developing empirically-based lie-detection procedures to be used by practitioners (e.g., the police) in applied settings. A number of new topics are just starting to be examined. These novel research avenues will surely yield interesting new findings in the future.

Keywords: Deception, lie detection, interviewing, CBCA, reality monitoring, polygraph.

Resumen

Detección de mentiras: estado de la cuestión y perspectivas de futuro.

Antecedentes: la detección de mentiras ha interesado a la humanidad a lo largo de la historia. También a los científicos, quienes han explorado diferencias psicológicas y conductuales al mentir vs. decir la verdad, así como modos de aumentar la precisión de la detección. **Método:** recientemente se han hecho avances sustanciales en esta área. En el presente artículo se revisan algunos de ellos. **Resultados:** se describen (a) las investigaciones y teorías contemporáneas sobre cómo la gente (intenta) detecta(r) mentiras; (b) los avances en procedimientos estratégicos de entrevista para detectar mentiras; (c) los hallazgos de meta-análisis recientes sobre aproximaciones sistemáticas para la detección verbal del engaño; y (d) algunos aspectos importantes de la detección psicofisiológica de la mentira. Al final del artículo se esbozan algunas tendencias emergentes y necesidades de investigación de cara al futuro. **Conclusiones:** el área de investigación de la detección de mentiras ha experimentado grandes desarrollos en tiempos recientes. A menudo (aunque no siempre) se ha centrado en desarrollar procedimientos de detección de mentiras de base empírica para su utilización en contextos aplicados (p. ej., por la policía). Algunas vías de indagación novedosas están empezando a explorar temas nuevos y, seguramente, darán lugar a futuros hallazgos nuevos e interesantes.

Palabras clave: engaño, detección de mentiras, entrevistas, CBCA, control de la realidad, polígrafo.

A person who gives poison may be recognized. He does not answer questions, or they are evasive answers; he speaks nonsense, rubs the great toe along the ground, and shivers; his face is discolored; he rubs the roots of the hair with his fingers; and he tries by every means to leave the house...
(Ayur-Veda, about 900 B.C.)

The above citation, borrowed from Trovillo (1939, p. 849), shows that people's interest in detecting the deceptions of dangerous others is longstanding. Deception detection has also interested psychology and communication scholars, who have explored, among other topics, the liars' behavior, the detectors'

strategies, and how to improve detection accuracy. In recent years, substantial advances have been made in the field. The goal of the current article is to briefly summarize some of these contributions, thus providing an updated (though necessarily incomplete because of space limitations) description of the state of the art in deception research. In the final section, some avenues for future research are outlined.

How people (try to) detect deception

Deception and behavior

Besides attesting to the longstanding nature of humans' preoccupation with deceit, the Ayur-Veda citation above also shows that three millennia ago people already believed that behavioral cues reveal deception. This belief has persisted throughout history, not only among lay people worldwide (Global Deception Research Team, 2006) but also among scientists, who have spent several decades trying to identify valid behavioral cues to deception (see

Ekman, 2009; Vrij, 2008; Zuckerman, DePaulo, & Rosenthal, 1981). However, recent meta-analyses have indisputably revealed that people can hardly detect deception from the observation of behavior (Bond & DePaulo, 2006), that the connection between lying and nonverbal cues is weak, as well as under the influence of a host of moderator variables (DePaulo et al., 2003; Sporer & Schwandt, 2006, 2007), and that cue training to detect deception hardly improves accuracy (Hauch, Sporer, Michael, & Meissner, 2016).

In short, people are poor judges of veracity, and even though they strongly believe that behavioral cues reveal deception (and focus on such cues when trying to assess veracity; see Bond, Howard, Hutchison, & Masip, 2013; Hartwig & Bond, 2011), meta-analytical evidence questions the utility of behavior as a source of deception markers.

Judgmental biases and the adaptive lie detector theory

A well-established finding in deception research is that lay people display a truth bias—that is, they tend to believe others are telling the truth rather than lying (Bond & DePaulo, 2006; Levine, Park, & McCornack, 1999). This bias is reduced or even reversed among practitioners whose job involves judging someone else's veracity, such as police officers (Masip, Alonso, Herrero, & Garrido, 2016; Meissner & Kassin, 2002).

The Adaptive Lie Detector Theory (ALIED; Street, 2015), which aims at explaining how people judge veracity, can help understand these divergent tendencies. Because behavioral deception cues are weak, the senders' statements often contain little or no information indicative of veracity. According to ALIED, under those circumstances, people make an informed guess based on context-general information. The base rate of truthful or deceptive statements is a kind of context-general information. Most statements that lay people regularly encounter in their daily lives are truthful; therefore, when they are uncertain about the veracity of a specific statement they tend to make truth judgments. However, practitioners such as police officers encounter deceptive messages more often than lay people; therefore, they are less likely to assume truthfulness when they are uncertain.

A major conceptual contribution of ALIED is that the common view that truth- and lie-biases are irrational tendencies that limit judgmental accuracy is replaced with the alternative notion that receivers with no access to specific diagnostic cues make the rational decision to focus on the general context to make the best possible guess. ALIED has been empirically supported in experimental research where cue diagnosticity has been manipulated. The findings revealed that the less diagnostic the cues, the more the participants used context-general information (specifically, the base-rates of lying) to assess veracity (Street, Bischof, Vadillo, & Kingstone, 2016).

Lie detection outside the laboratory

The above findings about people's poor lie-detection skills are mostly derived from laboratory experiments. In such experiments, observers are requested to make *immediate* judgments about the veracity of *unacquainted* senders' statements on the basis of *behavioral information* alone (which, as explained above, is poorly diagnostic of veracity). All these three elements make the task extremely challenging. Park, Levine, McCornack, Morrison,

and Ferrara (2002) asked participants (college students) to recall a lie they had detected in the past and to describe how they had detected it. They found that outside the laboratory lies are typically detected from *contextual* rather than behavioral information. Contextual information involves aspects such as physical evidence, third-party information, the liar's confession, and inconsistencies with prior knowledge. Furthermore, Park et al. found that outside the laboratory lies are typically detected in *familiar others* and *long after* they have been told. It is therefore apparent that the low accuracy rates derived from laboratory experiments may not be generalized to real-life contexts.

The superiority of contextual information compared to behavioral cues when it comes to judging veracity has also been demonstrated in experimental research. Both Blair, Levine, and Shaw (2010) and Bond et al. (2013) found in a series of experiments that observers reached higher accuracy rates when contextual information was available to them than when they had to base their veracity judgments on behavioral cues only.

Park et al.'s (2002) finding that in real life lies are typically detected from contextual information was replicated by Masip and Herrero (2015) with both police officers and community members. They also found that the very same participants who reported having detected lies from contextual (rather than behavioral) information in the past listed a number of behavioral cues when asked to indicate "how lies can be detected". This finding suggests that the allure of behavioral cues is strong when it comes to judging veracity. Indeed, Bond et al. (2013) demonstrated that people may forego perfectly diagnostic contextual information to base their judgments on poorly diagnostic behavioral cues.

We can therefore speculate that in real life people also focus on behavioral cues when they try to detect deception. However, this strategy is futile. In contrast, contextual information, either actively searched for by a persistent suspicious receiver or accidentally stumbled upon by a candid one, is indeed much more revealing.

The so-called situational familiarity effect looks consistent with the notion that contextual information is a better indicator of truth or deception than behavioral cues. Indeed, the veracity judgments of receivers who are familiar with the situation are more accurate than those of receivers unfamiliar with the situation. Presumably, the former compare the sender's statement with their situational knowledge to assess plausibility (Reinhard, Sporer, Scharmach, & Marksteiner, 2011). However, Reinhard, Scharmach, and Sporer (2012) found that *perceived* (not necessarily actual) familiarity is enough for the effect to occur. Therefore, the situational familiarity effect is caused, at least in part, by factors other than the validity of contextual cues.

Truth-default Theory

Recently, Levine (2014) proposed the Truth-default Theory (TDT). Rather than a unitary theory, TDT is a compilation of interrelated and logically coherent notions based on previous research. TDT provides a valuable framework to understand everyday life's deception and its detection. TDT's propositions, which are supported by empirical research (see Levine, 2014), are summarized in Table 1. Some of the notions expressed above (human's poor accuracy in judging veracity, their truth bias, the increased diagnostic value of contextual information relative to behavioral cues...) are incorporated into TDT. Propositions 13 and 14 are related to the contents in the next section.

Table 1
Propositions of Levine's (2014) Truth-default Theory

01. Most people tell the truth most of the time.
02. Most lies are told by a few prolific liars.
03. Most people believe what others say most of the time (truth bias).
04. This is adaptive (because most communications people encounter are honest) and enables efficient communication. However, it makes people vulnerable to occasional deception.
05. Both truthful and deceptive messages are means to attain certain goals. Most people do not lie if their goals can be attained telling the truth.
06. When the truth is inconsistent with the sender's goals, people may doubt veracity.
07. Other "triggers" raising suspicion are a lack of coherence (internal logical consistency) in message content, discrepancies between the message and the known reality, and third-party information revealing deception.
08. If these triggers are strong enough, the person will scrutinize the message to assess veracity.
09. The person may judge the message as deceptive on the basis of communication context and motive, sender demeanor, third-party information, and degree of coherence and correspondence.
10. Deception triggers may not occur at the time of the deception.
11. Because (except for a few transparent liars) the relationship between veracity and behavior is poor, deception is not accurately detected by passively observing the senders' behavior at the time the lie is told.
12. Instead, whenever deception is detected, this occurs later in time via the liar's confession, external evidence, or correspondence.
13. Context-sensitive questioning of the sender can produce diagnostic information. The wrong questioning may hinder detection accuracy.
14. Deception detection expertise does not involve skill in passively detecting and interpreting behavior but in generating diagnostic information from senders.

How to detect deception

The evidence that behavioral cues to deception have little diagnostic value has led to a shift in deception research. Many researchers are no longer interested in chasing elusive deception cues spontaneously displayed by the liar—such cues are weak and volatile. Instead, researchers are interested in designing interview strategies oriented to produce behavioral differences between truth-tellers and liars. In recent years, a huge amount of work has been conducted towards this goal, mainly in the laboratories of Vrij (UK) and Granhag (Sweden). The focus of this research is applied, as its ultimate goal is to provide the law enforcement with lie detection tools to be used when questioning crime suspects (Vrij & Fisher, 2016; Vrij & Granhag, 2012; Vrij, Granhag, & Porter, 2010). Within this new perspective, the lie detector is no longer conceptualized as an *observer* who attentively but *passively* scrutinizes the sender's behavior, but as an *interviewer* who *actively* employs specific questioning approaches that can reveal deception.

These interview approaches need to be based on psychological differences between truth tellers and liars. For instance, Granhag, Hartwig, Mac Giolla, and Clemens (2015) argue that guilty suspects (liars) are unwilling to provide information to the police because this may expose them. Therefore, they will use *avoidance strategies* such as being elusive about their whereabouts, providing only vague details, or giving details the police cannot verify. When faced with the evidence, liars will use the *escape strategy* of denying that evidence. Conversely, innocent suspects (truth tellers) are eager to provide information because they feel this can help establish their innocence. When faced with some form of incriminating evidence, truth tellers will generally be willing to admit it because they believe "the truth will shine through" and that if they committed no crime they cannot be convicted because people get what they deserve (see Kassin & Norwick, 2004; Masip & Herrero, 2013).

Strategic Use of Evidence

These differences between truth tellers and liars can be exploited to detect deception. For example, when some kind of evidence is available interviewers can use the Strategic Use of Evidence (SUE) technique (e.g., Hartwig, Granhag, & Luke, 2014). Interviewers using the SUE technique question the suspect about their whereabouts while withholding the available incriminating evidence until the end of the interview—i.e., during the interview, the suspect is unaware of the evidence against him or her. Guilty suspects are expected to carefully avoid mentioning any potentially incriminating information, which will elicit statement-evidence inconsistencies. For instance, close-circuit-television footage shows the suspect was near the crime scene just before the crime occurred, but the suspect states s/he was somewhere else. Conversely, innocent suspects will feel that because they are innocent they have nothing to hide and nothing to fear, and will therefore be more honest and forthcoming. As a result, their statements will be more consistent with the evidence. A meta-analysis showed that the difference between liars and truth tellers in terms of statement-evidence inconsistencies was substantially larger when the SUE technique was used than when the evidence was disclosed early in the interview (Hartwig et al., 2014).

Verifiability approach

The guilty suspects' tendency to withhold information is also exploited in the verifiability approach (Nahari, Vrij, & Fisher, 2014a). Guilty suspects lying about their alibi, particularly if they are explicitly requested by the interviewer to be very detailed, may feel that if they provide little detail they may look deceptive. However, if they provide many details the police can check on these details and find out that the alibi is false. Liars may resolve this dilemma by providing unverifiable details. Conversely, truth tellers will provide more verifiable details than liars. Verifiable details involve the description of activities performed with or in

the presence of other people the police may question, or in an area where the suspect believes there are surveillance cameras. They also involve admitting having performed activities that are regularly electronically recorded (e.g., using the credit card). Research has supported the notion that liars provide fewer verifiable details than truth tellers (e.g., Nahari et al., 2014a). Interestingly, this approach is immune to countermeasures; even if liars are aware that they must provide verifiable details, only truth tellers are in a position to provide them. In fact, a study showed that instructing suspects to give verifiable details resulted in an increase of such details among truth tellers but not among liars (Nahari, Vrij, & Fisher, 2014b). Thus, the *explicit request* to include verifiable details in the account increases the difference (in terms of this kind of details) between liars and truth tellers, thus enhancing the differentiation power of this technique.

Cognitive load approaches

Liars and truth tellers can also differ in terms of cognitive effort. Vrij et al. (2010) argued that creating a lie might require more cognitive effort than just describing an episodic memory. Therefore, during an interview, the liars' cognitive load might be higher than the truth tellers'. If cognitive load is artificially increased further, this may result in liars showing visible signs of mental overload. Research has tested the impact of a number of cognitive-load-inducing strategies on both behavioral cues and detection accuracy. Such strategies involved asking interviewees to describe the events in the reverse (instead of chronological) order, conducting the interview in a foreign language, or asking interviewees to stare at the interviewer's eyes or to perform a secondary task during the interview (for an overview, see Vrij, Fisher, & Blank, 2017).

Two major reviews have been published recently on the effectiveness of such strategies, one focused on the cues elicited (Vrij, Fisher, Blank, Leal, & Mann, 2016) and the other one focused on detection accuracy (Vrij et al., 2017). In addition to explicit cognitive-load-inducing procedures, these reviews included two additional strategies: First, encouraging interviewees to say more. As argued above, liars will presumably be less willing than truth tellers to add details, and will have to invent such details, which is cognitively difficult. Second, asking unexpected questions. Liars prepare for the interview, but they can prepare the answers to only those questions they can anticipate. Inventing answers to unexpected questions is mentally taxing and may result in little detail, implausible information, and contradictions between the answers of different suspects interviewed separately (e.g., Vrij et al., 2016).

The cue review revealed that the percentage of cognitive cues that discriminated in the predicted direction when using a cognitive lie-detection approach (65% of the cues examined) was larger than the percentage of all kinds of cues that discriminated in either direction when using a "standard" interviewing approach (30%). More specifically, the cognitive approach elicited significantly more detail, plausibility and consistency cues than the "standard" approach (Vrij et al., 2016).

The accuracy meta-analysis revealed that accuracy in distinguishing between truths and lies was higher when using a cognitive approach (71% accuracy) than when using a "standard" approach (56% accuracy), both when humans made the veracity judgments and when the number of objective cues (e.g., number

of details) were entered as predictors in statistical analyses that classified the statements as either truthful or deceptive (e.g., discriminant analyses). Interestingly, humans in these studies were *not* informed about the cues they had to use to make their judgments; had they been informed, accuracy would probably have been even higher. Each of the three strategies (i.e., using cognitive-load-inducing procedures, asking interviewees to say more, and asking unexpected questions) boosted accuracy (Vrij, Fisher et al., 2017).

A number of concerns have been raised concerning cognitive lie-detection approaches. First, there are many circumstances where lying is not cognitively more taxing than truth telling (e.g., Blandón-Gitlin, López, Masip, & Fenn, in press; Burgoon, 2015; Sporer, 2016). Second, strong cognitive-load-inducing techniques can elicit visible indicators of overload not only among liars but also among truth tellers. The so-called TRI-Con (Time Restricted Integrity-Confirmation) interview addresses this issue. When using TRI-Con, interviewers prompt interviewees about the general topic of the forthcoming questions. However, the specific questions are not revealed until the time they are asked. Such prompts activate truthful memories in working memory, which facilitates truthful responding but makes it cognitively harder to deceive, as liars must inhibit the activated memory and replace it with a fabrication (Walczyk et al., 2012). Third, the limits of cognitive lie-detection approaches need to be explored. For example, these approaches may not work to detect lies about intentions (Fenn, McGuire, Langben, & Blandón-Gitlin, 2015) or with stigmatized groups of people (Fenn, Blandón-Gitlin, Pezdek, & Yoo, 2016). Finally, the theoretical background of these approaches is generally weak; models specifying the specific cognitive mechanisms and processes involved in lying, which would allow for more precise and nuanced predictions, are needed (Blandón-Gitlin, Fenn, Masip, & Yoo, 2014; Blandón-Gitlin et al., in press; for one such model, see Walczyk, Harris, Duck, & Mulay, 2014).

Systematic verbal lie detection approaches

Although behavioral cues are generally poor indicators of deception, meta-analyses show that verbal cues are more diagnostic than nonverbal cues (DePaulo et al., 2003; Hauch et al., 2016). Some systematic approaches have been developed to assess credibility from the verbal content of extended free-narrative statements, such as the reality monitoring (RM) approach (Sporer, 2004), and Criteria-based Content Analysis (CBCA; Steller & Köhnken, 1989). Both are based on the notion that the verbal descriptions of self-experienced events differ from those of imagined or invented events.

Reality Monitoring (RM)

According to the RM approach, relative to imagined or invented memories, actual autobiographical memories (and their verbal descriptions) contain more contextual (time, space...), sensory, and semantic information, as well as fewer references to cognitive processes at the time of encoding. Reviews show that accuracy rates in separating truths (i.e., descriptions of memories of self-experienced events) from lies (inventions) with the RM verbal criteria are typically within the 60%-to-70% range (Masip, Sporer, Garrido, & Herrero, 2005; Vrij, 2008).

Criteria-based Content Analysis (CBCA)

CBCA emerged in forensic settings in Germany to separate between true and false allegations of child sexual abuse (Undeutsch, 1989). It contains 19 credibility criteria (Table 2; see, e.g., Raskin & Esplin, 1991; Steller & Köhnken, 1989, for criteria descriptions). CBCA experts assume that the more the criteria the child's statement contains (or the stronger the criteria), the more likely it describes a self-experienced event (see Volbert & Steller, 2014, for underlying theoretical premises). However, the absence of criteria should not be interpreted as indicative of deception (e.g., Raskin & Esplin, 1991).

CBCA is to be used within a more general assessment procedure, called Statement Validity Assessment (SVA), which systematically considers a number of alternative reasons for the child's allegation. SVA contains a semi-structured interview protocol to collect the child's account, considers the potential influence of a number of variables (cognitive or language limitations, suggestibility, etc.) on statement quality, and considers other sorts of information besides statement quality to make the credibility judgment (e.g., Raskin &

Esplin, 1991). Several countries admit SVA/CBCA assessments in court in child sexual abuse cases.

Although CBCA was developed to assess the credibility of alleged child victims' statements of sexual abuse, research has explored its usefulness to differentiate between truthful and deceptive statements of adults in addition to children, witnesses and suspects in addition to victims, and events other than sexual abuse (see Table 5 in Hauch, Sporer, Masip, & Blandón-Gitlin, in press, for the characteristics of CBCA studies).

CBCA is a clinical assessment procedure rather than a standardized psychometric test. However, its reliability and validity are important if it is to be used in forensic practice (Hauch et al., in press). A meta-analysis on the inter-rater reliability of CBCA revealed that most criteria have sufficient to good reliability (although whether reliability is high enough for CBCA/SVA evidence to be admitted in court is open to discussion). However, as shown in Table 2, whereas reliability was consistently high for those criteria with straightforward definitions, it was poor for criteria with less clear definitions (e.g., Criteria 2 and 9). The latter criteria should be used with great caution. Proportion

Table 2
Inter-rater reliability and validity of CBCA criteria according to recent meta-analyses

CBCA Criterion	Inter-rater Reliability (Hauch et al., in press)					Validity (Amado et al., 2015, 2016)	
	Pearson's r^a	Proportion agreement ^{a,b}	Mdn for kappa ^c	Mdn for weighted kappa and ICC ^c	Mdn for Maxwell's Random Error ^c	d_{children}^a	d_{adults}^a
General Characteristics							
01. Logical structure	.69	.79	.38	.18	.77	0.47	0.48
02. Unstructured production	.46	.70	.32	.55	.40	0.40	0.53
03. Quantity of details	.73	.70	.50	.75	.40	0.77	0.55
Specific Contents							
04. Contextual embedding	.71	.68	.43	.49	.67	0.69	0.19
05. Descriptions of interactions	.65	.77	.41	.53	.62	0.44	0.27
06. Reproduction of conversation	.86	.77	.61	.68	.73	0.53	0.34
07. Unexpected complications	.64	.79	.39	.45	.67	0.29	0.25
Peculiarities of Contents							
08. Unusual details	.62	.73	.52	.67	.49	0.27	0.31
09. Superfluous details	.52	.71	.48	.46	.52	0.42	0.14
10. Accurately reported details misunderstood	.81	.93	.47	.67	.91	0.31	0.22
11. Related external associations	.67	.83	.27	.55	.59	0.28	0.26
12. Accounts of subjective mental state	.79	.73	.58	.62	.61	0.46	0.18
13. Attribution of perpetrator's mental state	.76	.81	.55	.58	.64	0.18	0.09
Motivation-Related Contents							
14. Spontaneous corrections	.60	.71	.47	.59	.50	0.20	0.16
15. Admitting lack of memory	.78	.70	.50	.60	.53	0.15	0.25
16. Raising doubts about one's own testimony	.73	.90	.32	.48	.91	0.19	0.20
17. Self-deprecation	.79	.86	.61	.79	.81	0.16	0.04
18. Pardoning the perpetrator	.72	.80	.56	.46	.85	0.23	-0.02
Offense-Specific Elements							
19. Details characteristic of the offense	.71	.75	.23	-- ^d	.73	1.25	0.28
<i>Average</i>	.70	--	.45	.55	.62	0.40	0.25
<i>Total CBCA score</i>	.90	.76	.65	-- ^d	.51	0.78	0.55

^a Weighted analyses; ^b Without $\geq .999$ values; ^c Unweighted analyses; ^d Values not reported because the number of comparisons was too small

agreement showed the highest reliability values because, unlike other reliability coefficients, it does not correct for chance agreement. Base rates (i.e., the relative presence of each criterion in statements) influenced criterion reliability (see Hauch et al.'s in-press report for detail).

All meta-analytical estimates were very heterogeneous. Moderator analyses for Pearson's r revealed that reliability was higher in field studies and quasi-experiments than in laboratory settings. Note, however, that in all kinds of studies in this meta-analysis raters had been carefully trained (background literature reading, lectures, examples, practice with discussion and feedback, sometimes homework...) for many hours (for half the studies reporting training duration, the training lasted more than 8.75 hs; average training duration was 23 hs, $SD = 40$). Conversely, in non-research field settings raters can differ greatly in terms of training. This suggests the current estimates represent the upper reliability limits that can be achieved (Hauch et al., in press). The same considerations can be made concerning validity.

Hauch et al. (in press) suggest that CBCA experts testifying in court should include reliability estimates in their reports. Also, if several blind experts code different sections of the case statements, interrater reliability for a single case can be calculated and reported in court by the expert called to testify (Hauch et al., in press).

Concerning the CBCA validity, two meta-analyses have been published recently on the topic, one focused on children's reports (Amado, Arce, & Fariña, 2015), and one on adults' reports (Amado, Arce, Fariña, & Vilariño, 2016). All criteria significantly differentiated between truthful and deceptive statements of children, though (a) most effect sizes were small according to Cohen's (1988) guidelines (Table 2), and (b) for twelve criteria, effect sizes were not generalizable because of low inter-rater reliability (see the original report for effect sizes corrected for criterion unreliability and the associated credibility intervals). Similarly, all CBCA criteria but self-deprecation and pardoning the perpetrator significantly differentiated between truthful and deceptive accounts of adult participants; however, except for the general characteristics criteria set, effect sizes were small (Table 2). For adults, effect sizes were not generalizable (see original report).

The effect size for the total CBCA score was larger for children's accounts than it was for adults' accounts. Further, among children, it was substantially larger for field ($d = 2.40$) than for experimental studies ($d = 0.50$). However, as argued by Hauch et al. (in press), sum scores are problematic because (a) they make sense only if the different criteria measure a unidimensional construct, (b) validity differs across criteria, and (c) under certain circumstances, some CBCA criteria should weight more strongly than others.

Amado et al. (2016) also found that the average effect size across criteria (for adult participants) was larger in field studies ($d = 0.34$), particularly if they focused on sexual abuse or intimate partner violence ($d = 0.67$), than in experiments ($d = 0.25$). Surprisingly, the average effect size was larger for witnessed than for self-experienced events.

Overall, the general-characteristics criteria set appears to be the most valid, and the motivational set the least valid. Criteria 4 and 19 seem to discriminate very well with children, but not with adults. In general terms, CBCA as a whole seems to work better with children than with adults. It should be stressed, however, that CBCA must be used within the SVA framework, and that intensive

training in clinical psychology and psychological assessment is essential to properly understand and to be able to use CBCA and SVA (see Hauch et al., in press).

RM and CBCA

Oberlander et al. (2016) meta-analyzed the validity of both RM and CBCA. Rather than looking at individual criteria, they focused on the final credibility judgments made on the basis of either sum scores, statistical decisions, or the rater's personal decision. The overall effect size was $g = 1.03$, which is large and was significant. Assuming equal sensitivity and specificity, it would result in 70% of truths and 70 % of lies correctly detected (Oberlander et al., 2016). RM showed higher validity ($g = 1.26$) than CBCA ($g = 0.97$), but the difference was not significant. The full set of CBCA criteria permitted better discrimination than incomplete sets. Effect sizes did not differ across field and laboratory studies or for self-experienced vs. observed events; these null effects are at odds with Amado et al.'s (2015, 2016) findings for CBCA.

Psychophysiological detection of deception

Attempts were also made from ancient times to detect deception from the suspect's physiological reactions. The Ayur-Veda citation above refers to shivering and pallor, and Trovillo (1939) explains how the Greek physician Erasistratus (300-250 B.C.) was able to find out that Prince Antiochus of Syria was secretly in love with his young stepmother Stratonice by feeling his pulse. Some old lie-detection methods were based on the assumption that lying elicits fear, and reflect some understanding of physiology. For instance, in ancient China and India, crime suspects were given rice to chew; if they could not spit it out they were considered guilty. This ordeal reflects the observation that high stress reduces salivation (Kleinmuntz & Szucko, 1984).

Psychophysiological lie detection received a push in the 1920s. There was at the time a climate of reform towards police professionalization in the US that involved the adoption of scientific methods and procedures by the law enforcement (see Leo, 2009). In this context, Marston, Larson, and Keeler made innovations to record the suspect's heart rate, blood pressure, respiration, and skin conductance during questioning to assess the suspect's truthfulness (Alder, 2007; Bunn, 2012; Lykken, 1998). This was the beginning of polygraphic lie detection. More recently, electroencephalography and functional magnetic resonance imaging (fMRI) have also been tested as lie-detection procedures (e.g., Verschuere, Ben-Shakhar, & Meijer, 2011).

Lie detection tests

The two lie detection tests that have received most attention are the Comparison Question Test (CQT) and the Concealed Information Test (CIT). The CQT is used by the law enforcement in several countries around the world. Conversely, the CIT is rarely used in applied settings except in Japan, where it is ordinarily employed by the police (Ogawa, Matsuda, Tsuneoka, & Verschuere, 2015).

During a CQT the examinee is asked a series of irrelevant (e.g., "Is today Tuesday?"), relevant (e.g., "Did you murder Miss Smith?") and comparison questions (e.g., "During the first 20 years of your life, did you ever hurt anyone?"). The examinee

is instructed to respond “no” to comparison questions, but because they are deliberately vague and remote, the examinee is uncertain about the truthfulness of that response. The examinee is told that evidence of deception when responding to comparison questions would suggest s/he is the kind of person who could have committed the crime under investigation (e.g., Vrij, 2008). Guilty examinees are assumed to be more concerned by relevant than by comparison questions; therefore, they are expected to display the strongest physiological responding just after the relevant questions. Conversely, innocent examinees are expected to be more concerned by—and, hence, to react more strongly to—comparison than to relevant questions (e.g., Raskin, 1989). Iacono and Lykken (1997) conducted mail surveys on (a) members of the Society for Psychophysiological Research (SPR), and (b) fellows of Division 1 (General Psychology) of the American Psychological Association (APA). Most respondents believed the CQT is *not* based on any scientifically sound psychological principle or theory.

The CIT differs in many respects from the CQT. During a CIT, the examinee is asked a series of multiple-choice questions (e.g., “Which was the weapon used to murder Miss Smith? Was it ...a knife? ...a gun? ...a sword? ...a baseball bat? ...an axe? ...an arrow?”). For each question, only one of the response options (which can be presented either verbally or in pictorial form) is correct. Only those examinees who have knowledge about the crime details will consistently show stronger physiological reactions to correct than to incorrect alternatives through (most of) the test. Of note, the CIT does *not* attempt to detect *deception*, but *concealed knowledge*. In fact, the CIT can be used by the police to uncover new information (Ogawa et al., 2015). Imagine a person is missing and the police believes she was murdered by her partner. The police could run a CQT asking the suspect about the location of the body. The response alternatives would all be plausible locations. After the test, the police could check whether the body is indeed at the location suggested by the test results.

Unlike the CQT, which is used only with peripheral measures, the CIT is also used with central or “brain” measures such as fMRI and event-related potentials (ERPs). The most studied ERP in deception research is called P300, and is a positive electroencephalographic wave that starts at about 300 ms after the onset of the stimulus eliciting it. P300 has been found to accompany the recognition of meaningful information (e.g., Iacono, 2015; Rosenfeld, 2011). The CIT is also used with so-called behavioral measures—that is, reaction time and errors in replying “yes” or “no”, typically by pressing a specific key, after stimulus presentation. The CIT variant normally used with central and behavioral measures contains target, probe and irrelevant stimuli. Target and probe stimuli are presented only rarely (about 15% of the time each), whereas irrelevant stimuli are presented quite often (about 70% of the time). Targets (e.g., a facial picture of Miss Smith that was shown in the TV news the day after she was found dead) are known by both guilty and innocent examinees. During the test, all examinees admit they know targets by pressing the “yes” button. Because targets are known to all respondents, they will elicit physiological responding regardless of guilt. Probes (e.g., a picture of the weapon used to kill Miss Smith) are known by guilty but not by innocent examinees. All examinees press the “no” button when a probe is presented. However, because probes are meaningful for guilty (but not for innocent) suspects, guilty suspects will show physiological reactions when probes are presented. Finally, the frequent irrelevant stimuli are meaningless

for all examinees regardless of guilt and will thus elicit no physiological responding. Note that for innocent examinees probes are the same as irrelevant as irrelevant stimuli, whereas for guilty examinees probes are the same as meaningful as target stimuli.

The pattern of physiological responding during a CIT can be caused by the orienting response (OR), which occurs when an organism encounters a novel and/or significant stimulus (MacKay-Brandt, 2011). The OR allows the organism to determine how to react. It increases skin conductance, decreases heart rate, interrupts respiration, and is thought to produce a P300 wave. Response inhibition (i.e., the suppression of the dominant truthful response) can also play a role during a CIT, as it is associated with a decrease in heart rate and respiration, an increase of the P300 amplitude, and the activation of certain brain areas recorded with fMRI (see reviews by Meijer, Verschuere, Gamer, Merckelbach, & Ben-Shakhar, 2016; Verschuere & Meijer, 2014). Most respondents of Iacono and Lykken’s (1997) survey believed the CIT is based on scientifically sound psychological principles or theories.

Accuracy

Meijer and Verschuere (2015) presented an overview of available reviews examining the accuracy of the polygraph with both the CQT and the CIT. The top half of Table 3 displays the range of average sensitivity (accuracy in detecting lies or concealed information) and specificity (accuracy in detecting truths) reported in the reviews considered by Meijer and Verschuere. Note that because of the way field studies are conducted, accuracy rates for CQT field studies might be inflated (see Bull et al., 2004; Iacono, 1995). Also, only two individual CIT field studies were available for inclusion in Meijer and Verschuere’s (2015) overview. Notwithstanding these issues, it is apparent from Table 3 that the CQT does a better job in detecting liars than truth tellers.

Study type and measure	Sensitivity	Specificity
CQT - Polygraph ^{a,b}		
Laboratory studies	74%-82%	60%-83%
Field studies	84%-89%	59%-75%
CIT - Polygraph ^a		
Laboratory studies	76%-88%	83%-97%
Field studies ^c	42%-76%	94%-98%
ERPs ^d		
Laboratory studies	68% (range: 7%-100%)	82% (range: 31%-100%)
fMRI ^e		
Laboratory studies	84% (range: 55%-100%)	81% (range: 33%-100%)

^a Data from Meijer and Verschuere (2015). ^b Because the CQT (unlike the CIT) allows for inconclusive test results, a 74% sensitivity rate for the CQT does not mean that 26% of liars were misclassified as truth tellers; specifically, the percentage of liars misclassified as truth tellers ranged between 7% and 8% (depending on the review) in laboratory studies, and between 1% and 13% in field studies, whereas, the percentage of truth tellers misclassified as liars ranged between 10% and 16% in laboratory studies, and between 12% and 19% in field studies (Meijer & Verschuere, 2015). ^c Only two individual studies. ^d Data from Terol et al. (2014). ^e Data from Ganis (2015)

Apparently, relevant questions are perceived as more threatening not only by guilty examinees but also by some innocent ones. Conversely, the CIT performs better with innocent than with guilty examinees. Apparently, some guilty individuals do not encode—or have forgotten at the time of the test—certain details about the event that are later probed during the test.

In a review of 16 CIT studies measuring ERPs, Terol, Álvarez, Melgar, and Manzanero (2014) found the average accuracy rates displayed in the corresponding rows in Table 3. Before concluding that sensitivity is comparatively small, it is important to keep in mind that the authors included in their review several studies where guilty participants successfully tried to beat the test. It appears from these figures that ERPs, which are central measures, yield classification rates very similar to those obtained with a polygraph (but see a recent experiment conducted by Langleben et al., 2016, that challenges this conclusion).

Ganis (2015) reviewed ten fMRI studies in which the brain areas activated during deception were mapped and then an attempt was made to identify individual liars and truth tellers on the basis of their activation in these areas. Again, as shown in the bottom rows in Table 3, classification rates obtained with this sophisticated, cutting-edge brain imaging technology do not seem superior to those obtained with the old-fashioned polygraph.

Finally, in a recent meta-analysis Suchotzki, Verschuere, Van Bockstaele, Ben-Shakhar, and Crombez (2017) found an effect size $d = 1.297$ for the reaction time (RT) difference between truths and lies using the CIT. After calculating Rosenthal and Rubin's (1982; see also Fritz, Morris, & Richler, 2012) binomial effect size display (which assumes equal sensitivity and specificity) it becomes apparent that RT measures would result in 77% of truths and lies correctly identified.

Differential activation is a continuum, and examiners use somewhat arbitrary cutoff points on that continuum to categorize examinees as a truth tellers or liars. The percentage of liars (or truth tellers) correctly identified depends on the specific location of the cutoff point on the continuum. To calculate accuracy independently of specific cutoff points, some researchers turned to Receiver Operating Characteristic (ROC) curves. A ROC curve graphically displays all possible combinations of true positives (sensitivity), true negatives (specificity), false positives (truth tellers misclassified as liars), and false negatives (liars misclassified as truth tellers). Accuracy can be represented as a single value, namely the area under the ROC curve (AUC). An AUC = .50 represents chance accuracy, whereas an AUC = 1.00 denotes perfect accuracy (see Swets, Dawes, & Monahan, 2000, for detailed and clear information about ROC curves). Table 4 displays average and median AUC values for different measures reported in several meta-analyses (see Meijer et al., 2016). It is clear that, as stressed by Meijer et al. (2016), figures are very similar for peripheral, central, and behavioral measures. The AUC for fMRI appears to be higher, but Meijer et al. noted that this value is estimated from only four studies with few participants, and that because none of the studies used cross validation this figure might be an overestimation.

Countermeasures

Properly trained examinees can beat the polygraph test by using either physical (e.g., pressing one's toes to the floor) or mental (e.g., performing mental calculations) countermeasures (see Honts, 2014,

for a recent review). Test sensitivity will decrease if examinees successfully increase their physiological responding to comparison questions (CQT) or irrelevant response alternatives (CIT), and/or if they successfully decrease their responding to relevant questions (CQT) or response alternatives (CIT). Countermeasure-detection techniques, such as movement sensors to be placed in the chair, have been developed by the polygraph industry, but research is lacking examining their effectiveness (Honts, 2014).

An argument for the replacement of the traditional polygraph (which measures peripheral responses) with ERPs or fMRI (which measure central nervous system activity) is that the latter measures are not amenable to conscious manipulation by the examinees (e.g., Iacono, 2015). This argument is fallacious. ERP studies have shown that examinees can learn to use specific strategies that decrease the test sensibility substantially (e.g., Rosenfeld, Soskins, Bosh, & Ryan, 2004). Rosenfeld's research group designed a new ERP-based lie-detection test to overcome this problem (Rosenfeld, Hu, Labkovsky, Meixner, & Winograd, 2013). Regarding fMRI, in a study conducted by Ganis, Rosenfeld, Meixner, Kievit, and Schendan (2011) sensitivity decreased from 100% to only 33% after the participants used a very simple physical countermeasure—although its use could be detected in the fMRI images.

Summary

The relative sensitivity and specificity of the polygraph depends on whether the examiner uses the CQT or the CIT. Central measures (which are normally employed with the CIT) do not permit better discrimination than either peripheral or behavioral measures, and are vulnerable to countermeasures.

Future prospects

Deception detection seems to be as much of a timely topic for the near future as it was back in the remote Ayur-Veda times. Space limitations made it impossible to discuss some emerging research topics that will presumably gain momentum in the near future.

Table 4 Accuracy using different tests and measures expressed as the area under the ROC curve (AUC)		
Study type and measure		AUC
CQT – Laboratory Studies	Polygraph (peripheral measures) ^a	<i>Mdn</i> = .85
CQT – Field Studies	Polygraph (peripheral measures) ^a	<i>Mdn</i> = .89
CIT – Laboratory Studies	Polygraph (peripheral measures) ^a	<i>Mdn</i> = .88
	Skin conductance responding ^b	<i>M</i> = .85
	Respiration ^b	<i>M</i> = .77
	Heart rate ^b	<i>M</i> = .74
	P300 (ERP) ^b	<i>M</i> = .88
	Reaction time ^b	<i>M</i> = .82
	fMRI ^b	<i>M</i> = .94

Note: *Mdn* = median; *M* = weighted mean
^a Data from the National Research Council (2003). ^b Data from Meijer et al. (2014, 2016)

Specifically, current concerns about international terrorism and airport security have led researchers to investigate how to detect lies about future intentions (Granhag & Mac Giolla, 2014). The contemporary widespread use of communication technologies has led to the recent study of the influence of communication medium (face-to-face, phone, email, instant messaging...) on how much, how, to whom, and about what people lie, as well as on deception cues and lie detection (e.g., Smith, Hancock, Reynolds, & Birnholtz, 2014). Technological developments also led researchers to examine whether linguistic deception cues can be identified with computers, but success has been limited (Hauch, Blandón-Gitlin, Masip, & Sporer, 2015). Computers can also integrate large amounts of information (e.g., scores on a number of verbal content criteria) to help humans judge credibility. For example, a procedure based on High Dimensional Visualization combining multidimensional scaling and virtual reality modelling has been quite successful in separating truthful from deceptive statements (e.g., Manzanero, Alemany, Recio, Vallet, & Aróztegui, 2015).

Recent theories such as ALIED and TDT make testable predictions and will surely spur research. TDT has empirical support but some of its propositions would benefit from replication, and new propositions can be added (Masip & Herrero, 2015; Van Swol, 2014). Little is known about how people (try to) detect lies outside laboratory settings; inspired by TDT and the evidence reviewed in the relevant section above, we recently started an ambitious research program along this line.

Research on cognitive approaches to detect deception would benefit from stronger theoretical bases. A promising theory is Walczyk et al.'s (2014) Activation-Decision-Construction-Action Theory, which also makes many testable predictions that will stimulate research (e.g., Masip, Blandón-Gitlin, de la Riva, & Herrero, 2016). Sporer (2016) also made valuable theoretical contributions. The boundary conditions within which novel interview approaches to detect deception work need to be explored (Fenn et al., 2015), including the liars' countermeasures (Luke, Hartwig, Shamash, & Granhag, 2016). Research has

only started to examine how well practitioners can learn to use these new techniques (Vrij, Leal, Mann, Vernham, & Brankaert, 2015). Also, the application of such approaches to settings other than investigative interviewing needs to be explored (see Harvey, Vrij, Leal, Lafferty, & Nahari, 2017; Ormerod & Dando, 2014, for applications to insurance claims and airport contexts, respectively).

Systematic verbal credibility assessment approaches may benefit from testing new criteria and from high-quality field studies (which are rare) focusing on the kinds of cases forensic experts are called to testify about. Similarly, there are almost no field studies on the CIT. However, uncertainty concerning ground truth in real cases has always hampered field studies on deception detection, and will presumably continue to do so. The Japanese field use of the Searching-CIT (that allows for material corroboration of the test results) might allow researchers to conduct sound CIT studies in field settings. On the other hand, there are a number of aspects in the Japanese use of the CIT that differ from laboratory research (see Ogawa et al., 2015). The use of the polygraph in real criminal cases can have serious consequences for suspects; therefore, there is urgent need to examine under controlled laboratory conditions the impact of these peculiar practices on the test results.

ERP and fMRI are relatively new approaches. Many issues remain unexplored, in particular concerning fMRI. Indeed, more countermeasure research is needed. Other brain-imaging technologies that might conceivably develop in the future will surely stimulate lie-detection research.

Finally, the cognitive and reasoning functioning of both people with mental health challenges and people with intellectual disability differs from that of other individuals. Surprisingly, little research has been conducted on the production and detection of their lies (see, e.g., Manzanero et al., 2015, for an exception).

To conclude, these are exciting times for deception researchers. Many potential new avenues of inquiry lie before us. Only time will tell where this applied area of scientific research will lead us in the future.

References

- Alder, K. (2007). *The lie detectors: The history of an American obsession*. Lincoln, NE: University of Nebraska Press.
- Amado, B. G., Arce, R., & Fariña, F. (2015). Undeutsch hypothesis and Criteria-Based Content Analysis: A meta-analytic review. *The European Journal of Psychology Applied to Legal Context*, 7, 3-12.
- Amado, B. G., Arce, R., Fariña, F., & Vilarinho, M. (2016). Criteria-Based Content Analysis (CBCA) reality criteria in adults: A meta-analytic review. *International Journal of Clinical and Health Psychology*, 16, 201-210.
- Blair, J. P., Levine, T. R., & Shaw, A. S. (2010). Content in context improves deception detection accuracy. *Human Communication Research*, 36, 423-442.
- Blandón-Gitlin, I., Fenn, E., Masip, J., & Yoo, A. (2014). Cognitive-load approaches to detect deception: Searching for cognitive mechanisms. *Trends in Cognitive Sciences*, 18, 441-444.
- Blandón-Gitlin, I., López, R. M., Masip, J., & Fenn, E. (in press). Cognición, emoción y mentira: implicaciones para detectar el engaño [Cognition, emotion, and lying: Implications to detect deception]. *Anuario de Psicología Jurídica*. doi:10.1016/j.apj.2017.02.004
- Bond, C. F., Jr., & DePaulo, B. M. (2006). Accuracy of deception judgments. *Personality and Social Psychology Review*, 10, 214-234.
- Bond, C. F., Jr., Howard, A. R., Hutchison, J. L., & Masip, J. (2013). Overlooking the obvious: Incentives to lie. *Basic and Applied Social Psychology*, 35, 212-221.
- Bull, R., Baron, H., Gudjonsson, G., Hampson, S., Rippon, G., & Vrij, A. (2004). *A review of the current scientific status and fields of application of polygraphic deception detection*. London, UK: British Psychological Society.
- Bunn, G. (2012). *The truth machine: A social history of the lie detector*. Baltimore, MD: Johns Hopkins University Press.
- Burgoon, J. K. (2015). When is deceptive message production more effortful than truth-telling? A baker's dozen of moderators. *Frontiers in Psychology*, 6, 1965.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.
- DePaulo, B. M., Lindsay, J. J., Malone, B. E., Muhlenbruck, L., Charlton, K., & Cooper, H. (2003). Cues to deception. *Psychological Bulletin*, 129, 74-118.
- Ekman, P. (2009). *Telling lies: Clues to deceit in the marketplace, politics, and marriage*. New York, NY: Norton.
- Fenn, E., Blandón-Gitlin, I., Pezdek, K., & Yoo, A. (2016). *Minority status matters: The potential role of stereotype threat in criminal investigations using a cognitive load approach*. Manuscript submitted for publication.

- Fenn, E., McGuire, M., Langben, S., & Blandón-Gitlin, I. (2015). A reverse order interview does not aid deception detection regarding intentions. *Frontiers in Psychology, 6*, 1298.
- Fritz, C. O., Morris, P. E., & Richler, J. J. (2012). Effect size estimates: Current use, calculations, and interpretation. *Journal of Experimental Psychology: General, 141*, 2-18.
- Ganis, G. (2015). Deception detection using neuroimaging. In P.-A. Granhag, A. Vrij & B. Verschuere (Eds.), *Detecting deception: Current challenges and cognitive approaches* (pp. 105-121). Chichester, UK: Wiley.
- Ganis, G., Rosenfeld, J. P., Meixner, J., Kievit, R. A., & Schendan, H. E. (2011). Lying in the scanner: Covert countermeasures disrupt deception detection by functional magnetic resonance imaging. *Neuroimage, 55*, 312-319.
- Global Deception Research Team (2006). A world of lies. *Journal of Cross-Cultural Psychology, 37*, 60-74.
- Granhag, P.-A., Hartwig, M., Mac Giolla, E., & Clemens, F. (2015). Suspects' verbal counter-interrogation strategies: Towards an integrative model. In P.-A. Granhag, A. Vrij & B. Verschuere (Eds.), *Detecting deception: Current challenges and cognitive approaches* (pp. 293-313). Chichester, UK: Wiley.
- Granhag, P.-A., & Mac Giolla, E. (2014). Preventing future crimes. Identifying markers of true and false intent. *European Psychologist, 19*, 195-206.
- Hartwig, M., & Bond, C. F., Jr. (2011). Why do lie-catchers fail? A lens model meta-analysis of human lie judgments. *Psychological Bulletin, 137*, 643-659.
- Hartwig, M., Granhag, P. A., & Luke, T. (2014). Strategic use of evidence during investigative interviews: The state of the science. In D. C. Raskin, C. R. Honts & J. C. Kircher (Eds.), *Credibility assessment: Scientific research and applications* (pp. 1-36). San Diego, CA: Academic Press.
- Harvey, A. C., Vrij, A., Leal, S., Lafferty, M., & Nahari, G. (2017). Insurance based lie detection: Enhancing the verifiability approach with a model statement component. *Acta Psychologica, 174*, 1-8.
- Hauch, V., Blandón-Gitlin, I., Masip, J., & Sporer, S. L. (2015). Are computers effective lie detectors? A meta-analysis of linguistic cues to deception. *Personality and Social Psychology Review, 19*, 307-342.
- Hauch, V., Sporer, S. L., Masip, J., & Blandón-Gitlin, I. (in press). Can credibility criteria be assessed reliably? A meta-analysis of Criteria-based Content Analysis. *Psychological Assessment*. doi:10.1037/pas0000426
- Hauch, V., Sporer, S. L., Michael, S. W., & Meissner, C. A. (2016). Does training improve detection of deception? A meta-analysis. *Communication Research, 43*, 283-343.
- Honts, C. R. (2014). Countermeasures and credibility assessment. In D. C. Raskin, C. R. Honts & J. C. Kircher (Eds.), *Credibility assessment: Scientific research and applications* (pp. 131-158). San Diego, CA: Academic Press.
- Iacono, W. G. (1995). Offender testimony: Detection of deception and guilty knowledge. In N. Brewer & C. Wilson (Eds.), *Psychology and policing* (pp. 155-171). Hillsdale, NJ: Lawrence Erlbaum.
- Iacono, W. G. (2015). Forensic application of event-related potentials to detect guilty knowledge. In P.-A. Granhag, A. Vrij & B. Verschuere (Eds.), *Detecting deception: Current challenges and cognitive approaches* (pp. 81-103). Chichester, UK: Wiley.
- Iacono, W. G., & Lykken, D. T. (1997). The validity of the lie detector: Two surveys of scientific opinion. *Journal of Applied Psychology, 82*, 426-433.
- Kassin, S. M., & Norwick, R. J. (2004). Why suspects waive their Miranda rights: The power of innocence. *Law and Human Behavior, 28*, 211-221.
- Kleinmuntz, B., & Szucko, J. J. (1984). Lie detection in ancient and modern times. *American Psychologist, 39*, 766-776.
- Langleben, D. D., Hakun, J. G., Seelig, D., Wang, A.-L., Ruparel, K., Bilker, W. B., & Gur, R. C. (2016). Polygraph and functional magnetic resonance imaging in lie detection: A controlled blind comparison using the concealed information test. *Journal of Clinical Psychiatry, 77*, 1372-1380.
- Leo, R. (2009). *Police interrogation and American justice*. Cambridge, MA: Harvard University Press.
- Levine, T. R. (2014). Truth-default theory (TDT). A theory of human deception and deception detection. *Journal of Language and Social Psychology, 33*, 378-392.
- Levine, T. R., Park, H. S., & McCornack, S. A. (1999). Accuracy in detecting truths and lies: Documenting the "veracity effect". *Communication Monographs, 66*, 125-144.
- Luke, T. J., Hartwig, M., Shamash, B., & Granhag, P.-A. (2016). Countermeasures against the Strategic Use of Evidence technique: Effects on suspects' strategies. *Journal of Investigative Psychology and Offender Profiling, 13*, 131-147.
- Lykken, D. T. (1998). *A tremor in the blood. Uses and abuses of the lie detector*. New York, NY: Plenum Press.
- MacKay-Brandt, A. (2011). Orienting response. In J. S. Kreutzer, J. DeLuca & B. Caplan (Eds.), *Encyclopedia of clinical psychology* (pp. 1830-1831). New York, NY: Springer.
- Manzanero, A. L., Alemany, A., Recio, M., Vallet, R., & Aróztegui, J. (2015). Evaluating the credibility of statements given by persons with intellectual disability. *Anales de Psicología, 31*, 338-344.
- Masip, J., Alonso, H., Herrero, C., & Garrido, E. (2016). Experienced and novice officers' Generalized Communication Suspicion and veracity judgments. *Law and Human Behavior, 40*, 169-181.
- Masip, J., Blandón-Gitlin, I., de la Riva, C., & Herrero, C. (2016). An empirical test of the decision to lie component of the Activation-Decision-Construction-Action Theory (ADCAT). *Acta Psychologica, 169*, 45-55.
- Masip, J., & Herrero, C. (2013). "What would you say if you were guilty?" Suspects' strategies during a hypothetical Behavior Analysis Interview concerning a serious crime. *Applied Cognitive Psychology, 27*, 60-70.
- Masip, J., & Herrero, C. (2015). Police detection of deception: Beliefs about behavioral cues to deception are strong even though contextual evidence is more useful. *Journal of Communication, 65*, 125-145.
- Masip, J., Sporer, S. L., Garrido, E., & Herrero, C. (2005). The detection of deception with the Reality Monitoring approach: A review of the empirical evidence. *Psychology, Crime, & Law, 11*, 99-122.
- Meijer, E. H., & Verschuere, B. (2015). The polygraph: Current practice and new approaches. In P.-A. Granhag, A. Vrij & B. Verschuere (Eds.), *Detecting deception: Current challenges and cognitive approaches* (pp. 59-80). Chichester, UK: Wiley.
- Meijer, E. H., Verschuere, B., Gamer, M., Merckelbach, H., & Ben-Shakhar, G. (2016). Deception detection with behavioral, autonomic, and neural measures: Conceptual and methodological considerations that warrant modesty. *Psychophysiology, 53*, 593-604.
- Meissner, C. A., & Kassin, S. M. (2002). "He's guilty!": Investigator bias in judgments of truth and deception. *Law and Human Behavior, 26*, 469-480.
- Nahari, G., Vrij, A., & Fisher, R. P. (2014a). Exploiting liars' verbal strategies by examining the verifiability of details. *Legal and Criminological Psychology, 19*, 227-239.
- Nahari, G., Vrij, A., & Fisher, R. P. (2014b). The verifiability approach: Countermeasures facilitate its ability to discriminate between truths and lies. *Applied Cognitive Psychology, 28*, 122-128.
- National Research Council. Committee to Review the Scientific Evidence on the Polygraph. Division of Behavioral and Social Sciences and Education (2003). *The polygraph and lie detection*. Washington, DC: National Academies Press.
- Oberlander, V., Naefgen, C., Koppehele-Gossel, J., Quinten, L., Banse, R., & Schmidt, A. F. (2016). Validity of content-based techniques to distinguish true and fabricated statements: A meta-analysis. *Law and Human Behavior, 40*, 440-457.
- Ogawa, T., Matsuda, I., Tsuneoka, M., & Verschuere, B. (2015). The Concealed Information Test in the laboratory versus Japanese field practice: Bridging the scientist-practitioner gap. *Archives of Forensic Psychology, 1*(2), 16-27.
- Ormerod, T. C., & Dando, C. (2014). Finding a needle in a haystack: Toward a psychologically informed method for aviation security screening. *Journal of Experimental Psychology: General, 144*, 76-84.
- Park, H. S., Levine, T. R., McCornack, S. A., Morrison, K., & Ferrara, S. (2002). How people really detect lies. *Communication Monographs, 69*, 144-157.
- Raskin, D. C. (1989). Polygraph techniques for the detection of deception. In D. C. Raskin (Ed.), *Psychological methods in criminal investigation and evidence* (pp. 247-296). New York, NY: Springer.
- Raskin, D. C., & Esplin, P. W. (1991). Statement Validity Assessment: Interview procedures and content-analysis of children's statements of sexual abuse. *Behavioral Assessment, 13*, 265-291.

- Reinhard, M.-A., Scharmach, M., & Sporer, S. L. (2012). Situational familiarity, efficacy expectations, and the process of credibility attribution. *Basic and Applied Social Psychology*, *34*, 107-127.
- Reinhard, M.-A., Sporer, S. L., Scharmach, M., & Marksteiner, T. (2011). Listening, not watching: Situational familiarity and the ability to detect deception. *Journal of Personality and Social Psychology*, *101*, 467-484.
- Rosenfeld, J. P. (2011). P300 in detecting concealed information. In B. Verschuere, G. Ben-Shakhar & E. Meijer (Eds.), *Memory detection: Theory and application of the Concealed Information Test* (pp. 63-89). Cambridge, UK: Cambridge University Press.
- Rosenfeld, J. P., Hu, X., Labkovsky, E., Meixner, J., & Winograd, M. R. (2013). Review of recent studies and issues regarding the P300-based complex trial protocol for detection of concealed information. *International Journal of Psychophysiology*, *90*, 118-134.
- Rosenfeld, J. P., Soskins, M., Bosh, G., & Ryan, A. (2004). Simple, effective countermeasures to P300-based tests of detection of concealed information. *Psychophysiology*, *41*, 205-219.
- Rosenthal, R., & Rubin, D. B. (1982). A simple, general purpose display of magnitude of experimental effect. *Journal of Educational Psychology*, *74*, 166-169.
- Smith, M. E., Hancock, J. T., Reynolds, L., & Birnholtz, J. (2014). Everyday deception or a few prolific liars? The prevalence of lies in text messaging. *Computers in Human Behavior*, *41*, 220-227.
- Sporer, S. L. (2004). Reality monitoring and the detection of deception. In P.-A. Granhag & L. A. Strömwall (Eds.), *The detection of deception in forensic contexts* (pp. 64-102). Cambridge, UK: Cambridge University Press.
- Sporer, S. L. (2016). Deception and cognitive load: Expanding our horizon with a working memory model. *Frontiers in Psychology*, *7*, 420.
- Sporer, S. L., & Schwandt, B. (2006). Paraverbal indicators of deception: A meta-analytic synthesis. *Applied Cognitive Psychology*, *20*, 421-446.
- Sporer, S. L., & Schwandt, B. (2007). Moderators of nonverbal indicators of deception: A meta-analytic synthesis. *Psychology, Public Policy, and Law*, *13*, 1-34.
- Steller, M., & Köhnken, G. (1989). Criteria-based Statement Analysis. In D. C. Raskin (Ed.), *Psychological methods in criminal investigation and evidence* (pp. 217-245). New York, NY: Springer.
- Street, C. N. H. (2015). ALIED: Humans as adaptive lie detectors. *Journal of Applied Research in Memory and Cognition*, *4*, 335-343.
- Street, C. N. H., Bischof, W. F., Vadillo, M. A., & Kingstone, A. (2016). Inferring others' hidden thoughts: Smart guesses in a low diagnostic world. *Journal of Behavioral Decision Making*, *29*, 539-549.
- Suchotzki, K., Verschuere, B., Van Bockstaele, B., Ben-Shakhar, G., & Crombez, G. (2017). Lying takes time: A meta-analysis on reaction time measures of deception. *Psychological Bulletin*, *143*(4), 428-453.
- Swets, J. A., Dawes, R. M., & Monahan, J. (2000). Psychological science can improve diagnostic decisions. *Psychological Science in the Public Interest*, *1*, 1-26.
- Terol, O., Álvarez, M., Melgar, N., & Manzanero, A. (2014). Detección de información oculta mediante potenciales relacionados con eventos [Detection of concealed information using event-related potentials]. *Anuario de Psicología Jurídica*, *24*, 49-55.
- Trovillo, P. V. (1939). A history of lie detection. *The Journal of Criminal Law and Criminology*, *29*, 848-881.
- Undeutsch, U. (1989). The development of Statement Reality Analysis. In J. C. Yuille (Ed.), *Credibility assessment* (pp. 101-133). Dordrecht, Holland: Kluwer.
- Van Swol, L. M. (2014). Questioning the assumptions of deception research. *Journal of Language and Social Psychology*, *33*, 411-416.
- Verschuere, B., Ben-Shakhar, G., & Meijer, E. (Eds.) (2011). *Memory detection: Theory and application of the Concealed Information Test*. Cambridge, UK: Cambridge University Press.
- Verschuere, B., & Meijer, E. (2014). What's on your mind? Recent advances in memory detection using the Concealed Information Test. *European Psychologist*, *19*, 162-171.
- Volbert, R., & Steller, M. (2014). Is this testimony truthful, fabricated, or based on false memory? Credibility assessment 25 years after Steller and Köhnken (1989). *European Psychologist*, *19*, 207-220.
- Vrij, A. (2008). *Detecting lies and deceit. Pitfalls and opportunities*. Chichester, UK: Wiley.
- Vrij, A., & Fisher, R. (2016). Which lie detection tools are ready for use in the criminal justice system? *Journal of Research in Memory and Cognition*, *5*, 302-307.
- Vrij, A., Fisher, R., & Blank, H. (2017). A cognitive approach to lie detection: A meta-analysis. *Legal and Criminological Psychology*, *22*, 1-21.
- Vrij, A., Fisher, R., Blank, H., Leal, S., & Mann, S. (2016). A cognitive approach to elicit verbal and nonverbal cues to deceit. In J. W. Van Prooijen, & P. A. M. Van Lange (Eds.), *Cheating, corruption, and concealment* (pp. 284-310). Cambridge, UK: Cambridge University Press.
- Vrij, A., & Granhag, P.-A. (2012). Eliciting cues to deception and truth: What matters are the question asked. *Journal of Applied Research in Memory and Cognition*, *1*, 110-117.
- Vrij, A., Granhag, P.-A., & Porter, S. (2010). Pitfalls and opportunities in nonverbal and verbal lie detection. *Psychological Science in the Public Interest*, *11*, 89-121.
- Vrij, A., Leal, S., Mann, S., Vernham, Z., & Brankaert, F. (2015). Translating theory into practice: Evaluating a cognitive lie detection training workshop. *Journal of Applied Research in Memory and Cognition*, *4*, 110-120.
- Walezyk, J. J., Griffith, D. A., Yates, R., Visconte, S. R., Simoneaux, B., & Harris, L. L. (2012). Lie detection by inducing cognitive load. Eye movements and other cues to the false answers of "witnesses" to crimes. *Criminal Justice and Behavior*, *39*, 887-909.
- Walezyk, J. J., Harris, L. L., Duck, T. K., & Mulay, F. (2014). A social-cognitive framework for understanding serious lies: Activation-decision-construction- action theory. *New Ideas in Psychology*, *34*, 22-36.
- Zuckerman, M., DePaulo, B. M., & Rosenthal, R. (1981). Verbal and nonverbal communication of deception. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (vol. 14, pp. 1-60). New York, NY: Academic Press.