Psychology, Crime & Law
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/gpcl20

Effectiveness of deception detection training: a meta-analysis

James E. Driskell [a]

[a] Florida Maxima Corporation, Winter Park, FL, USA

Available online: 27 Jun 2011

Psychology, Crime & Law
2011, 1–19, iFirst

Effectiveness of deception detection training: a meta-analysis

James E. Driskell*

Florida Maxima Corporation, Winter Park, FL, USA

(*Received 19 May 2010; final version received 21 October 2010*)

A meta-analysis was conducted to determine the overall effectiveness of deception detection training and to identify conditions that may moderate training effectiveness. The analysis was based on a total of 16 studies with 30 separate hypothesis tests, representing the behavior of 2847 trainees. Results indicated that the effect of deception detection training on detection accuracy was positive, significant, and of medium magnitude. Moreover, training effectiveness was moderated by the type of training implemented, training content, trainee expertise, and the type of lie told. Discussion centers on implications for training design and implementation.

Keywords: detection of deception; lie detection; training

Introduction

People attempt to deceive one another for a myriad of reasons, some trivial and some of significant practical import. Because the consequences of deception can be substantial, researchers have long been concerned with the practical task of detecting deception. Recent world events have led to an increased emphasis placed on the capability to detect deception, especially in applied settings such as security checkpoints or screening contexts in airports, bus terminals, or train stations. In these environments, perceivers attempt to detect deception on the basis of observable behavioral cues, such as facial expression, body movement, vocal cues, and patterns of speech. However, empirical evidence indicates that neither naive laypersons nor expert law enforcement personnel are able to discern truth from deception with any impressive measure of accuracy (Bond & DePaulo, 2006; DePaulo, 1994). In fact, Bond and DePaulo (2006) concluded that 'the average person discriminates lies from truth at a level slightly better than he or she would achieve by flipping a coin' (p. 230). Moreover, DePaulo (1994) noted that despite some reported success, 'I am not optimistic about the prospects of teaching these cues directly' (p. 85). The fact that we are not naturally adept at detecting deception is not in itself too surprising. However, from an applied psychology perspective, the claim that training is of little value in enhancing performance would be surprising, and runs counter to other empirical evidence regarding training effectiveness (Arthur, Bennett, Edens, & Bell, 2003; Keith & Frese, 2008). The purpose of this study is to integrate the literature on deception detection training, provide a summary of the overall effectiveness of

*Email: jdriskell@rollins.edu

ISSN 1068-316X print/ISSN 1477-2744

DOI: 10.1080/1068316X.2010.535820
http://www.informaworld.com

deception detection training, and specify the conditions under which deception detection training is most effective.

Deception detection training

In a typical study of deception detection training, stimuli materials are first created by having a set of participants produce truthful and deceptive statements. For example, research participants may commit some transgression and then either tell the truth or lie regarding that transgression. The interviews with these participants are videotaped, and a set of stimulus videotapes are generated that contain segments of truthful and deceptive behavior. In the training portion of the study, a new set of participants are selected whose task is to view these videotapes to discern whether the individual in each instance is telling the truth or lying. The *deception detection training* group receives training designed to enhance accuracy of detection. The *control* group receives no specialized training. Participants then view the stimulus videotapes, and data are recorded on the mean accuracy, or percentage of correct responses, for training and control groups.

Conclusions regarding whether this type of training is effective vary. In a narrative review of deception detection training research, Bull (2004) concluded that the effectiveness of training was minimal. Akehurst, Bull, Vrij, and Kohnken (2004) noted that 'There are very mixed findings with regard to the trainability of lie detection skills ... with little evidence that training actually works' (p. 878). On the other hand, some claim that the effects of deception detection training are positive. Frank and Feeley (2003) integrated the results of 11 studies and reported a modest effect of deception detection training, and Levine, Feeley, McCornack, Hughes, and Harms (2005) concluded that deception detection training results in a slight to moderate increase in accuracy.

Therefore, one goal of the present effort is to provide a precise empirical summary of the weight of available evidence on the efficacy of deception detection training. It should be noted that Frank and Feeley (2003) have conducted a prior meta-analytic integration of research on lie detection training. This integration incorporated the results of 11 studies, with $k = 20$ separate hypothesis tests, and reported that training resulted in a small gain in detection accuracy. Frank and Feeley's report represents a useful and timely effort to integrate this literature. However, one limitation of the Frank and Feeley analysis is that their goal was solely to determine the aggregate effect size for deception detection training. Although information regarding the overall efficacy of training is valuable, this information does not allow us to examine why certain training interventions are more effective than others. This issue is especially relevant given the significant variability of the effect sizes reported by Frank and Feeley (2003).

Meta-analysis has two important but complementary functions: synthesis and analysis (Driskell & Mullen, 2005; Rosenthal, Hiller, Bornstein, Berry, & Brunell-Neuleib, 2001). The synthesis function involves questions of central tendency and variability of a collection of effect sizes as a whole. The analysis function involves examining variables that may account for the variability within the collection of effect sizes. In fact, this capacity for meta-analysis to account for systematic variability in effect sizes, and to render precise tests of the effects of theoretically relevant and practically important moderators, can be one of the greatest contributions of a meta-analysis to the understanding of a phenomenon (Mullen, 1989). Therefore, a second goal of this analysis is to examine the extent to which the efficacy of deception detection training increases or decreases as a function of certain theoretically relevant and practically important moderators. These moderators are described in the following.

Effects of moderators

Type of training

Experts note that a comprehensive training program incorporates the presentation of requisite knowledge or information, the opportunity to practice skills, and feedback on correct and incorrect performance (Driskell, Salas, Johnston, & Wollert, 2008; Tannenbaum & Yukl, 1992). Indeed, Crews et al. (2007) noted that 'explicit instruction, practice, and feedback are three critical components of deception detection training' (p. 33). Although the 'typical' study was described in previous paragraphs, in practice, different studies often implement different types of training. Within this database, some deception training studies only provided trainees with information regarding cues to deception prior to testing (e.g. Kassin & Fong, 1999), some training studies only provided feedback regarding correct responses (e.g. Zuckerman, Koestner, & Alton, 1984), and some training studies provided a combination of information, practice, and feedback (e.g. deTurck, Feeley, & Roman, 1997). To address the question of what type of training is most beneficial, each hypothesis test was coded as to whether the training intervention incorporated information only, feedback only, or a combination of information, practice, and feedback.

Training content

The belief in the value of deception detection training is based on the assumption that there are reliable cues that distinguish between truth and lies, and that training is implemented that incorporates these cues. A number of authors have noted that that value of deception detection training is contingent on reliable training content, and have expressed concern that training may in some cases be counterproductive to the extent that it focuses attention on cues that are themselves not diagnostic in discerning deception (Kassin & Fong, 1999). One way to address this issue is to examine the extent to which each training study incorporated cues that have been shown to be reliable indicators of deception. In a comprehensive meta-analysis, DePaulo et al. (2003) identified 24 specific cues within five broad dimensions or categories that were significant predictors of deception. Drawing on these results, we first reviewed the 'training description' section of each study in the current database and listed those cues that were described as comprising each training intervention. (Note that this training information was only available for those studies that incorporated information on cues to deception as part of the training intervention; thus studies that included feedback-only interventions were excluded from this analysis.) We then selected from the set of 24 significant cues to deception reported in DePaulo et al. only those cues that were present in at least three of the studies in our database. This resulted in the identification of seven specific cues from within five categories of predictors of deception, as shown in Table 1. To provide an example: DePaulo et al. provide one subcategory of cues to deception described as 'Liars make less sense than truth tellers' that includes three specific cues that were shown to be significant predictors of deception. Two of these cues are represented in at least three of the deception detection training studies: 'communications are less logical' and 'communications are more discrepant.' We then coded each hypothesis test for whether or not the training included a cue from within this category (that is, the instructions to trainees either referenced the cue 'less logical' or the cue 'more discrepant'). In this manner, we coded each hypothesis test for whether or not it included cues within each of the five categories shown in Table 1. This strategy allows us to examine what specific training content best predicts training outcomes. Or, more specifically, given that these cues have been identified in previous research as significant cues to deception, we are able to examine whether training that incorporates these cues is more effective.[1]

Motivation of the deceiver

As Levine et al. (2005) and DePaulo et al. (2003) have noted, cues to deception are more likely to be evident when speakers are motivated to deceive than when there is little or no consequence attached to their performance. That is, speakers who are attempting to deceive are more likely to show signs of tension when there are consequences for success or failure. Accordingly, we would expect training to be more successful in those situations in which deceivers are motivated to lie (and thus the observable cues to deception are ostensibly more evident). Adopting the procedure reported in DePaulo et al. (2003), studies were coded into two categories; those that reported some type of identity-relevant or instrumental incentive for deceivers and studies that reported no incentive. Although we can distinguish between some motivation and no motivation in this manner, it is important to note that the level of motivation that is typically present in these research studies is substantially different than the level of motivation in a real-life, high stakes situation, a point that will be addressed more fully in the discussion.

Table 1. Deception categories and specific cues to deception.

Categories	Specific cues
Liars are more tense than truth tellers	More fidgeting
Liars are less forthcoming than truth tellers	Shorter speaking duration Fewer details
Liars make less sense than truth tellers	Less logical More discrepant/ambivalent
Liars are less engaged or expressive than truth tellers	Fewer illustrators
Liars are less fluent than truth tellers	More speech disturbances/word and phrase repetitions

Trainee expertise

Some have argued that training may benefit those who have experience more than naive persons because training can build on an existing scaffold of past experience (Kassin & Fong, 1999). On the other hand, others have noted that those with special experience or background in detection of deception may have little advantage over laypersons (DePaulo, 1994) or in fact may be less accurate because of existing biases (Meissner & Kassin, 2002). To examine whether trainee expertise moderates training effectiveness, each hypothesis test was coded as to whether the trainee population was naive (those, typically students, with no special background in detection of deception) or possessed special expertise in the detection of deception (typically law enforcement personnel).

Type of lie

Lies can either be about transgressions, in which the deception is to cover-up what one did, or about non-transgressions, in which the deception is to cover-up one's feelings or opinions. DePaulo et al. (2003) found that cues to deception tended to be stronger when lies were about transgressions than about feelings or opinions. Accordingly, training may be more successful when the deceiver lies about transgressions (and ostensibly the cues to deception are stronger) than when the deceiver lies about feelings or opinions. To examine whether the type of lie moderates training effectiveness, each hypothesis test was coded as to whether the lie told was regarding a transgression or was not about a transgression.

Length of training

Some studies of deception detection training report very brief training interventions of 15 minutes or less, and some report longer training sessions of several hours duration. In general, we would expect that more training would lead to greater results (see Driskell, Willis, & Copper, 1992), and Frank and Feeley (2003) recommended that, to be effective, deception detection training should last more than one hour. On the other hand, deTurck and Miller (1990) have cautioned that too much practice may lead to boredom and inhibit detection accuracy. To examine whether the effectiveness of deception detection training is related to how much training is provided, we recorded the length of training (in minutes) reported for each hypothesis test.

Procedure

Using all of the standard literature search techniques, an exhaustive search was conducted for studies examining deception detection training. Specifically, online computer searches were conducted on PsycNET, using the keywords deception, lying, and training. These computer searches were supplemented by ancestry approach searches (i.e. locating previous studies identified in the reference sections of already-located studies) and descendency approach searches (i.e. locating subsequent studies identified in Social Science Citation Index as having cited already-located studies), and scanning the past 25 years of leading psychology and social science journals (see Mullen, 1989, for a discussion of literature search techniques). It should be emphasized that any available previous reviews (such as Bull, 2004; DePaulo, 1994; Frank & Feeley, 2003; inter alia) were carefully scrutinized for potentially includable studies.

In order to be included in the present effort, a study had to provide one or more clear and unequivocal tests of the effect of deception detection training on judgment accuracy. Studies were included if they met the following criteria. First, deception detection training was defined operationally as a training intervention designed to enhance deception judgment accuracy. Second, studies had to employ a deception detection training intervention incorporating either information, practice, and/or feedback. Further, to be included in the current analysis, a study must report (or allow the retrieval of) a test of the effectiveness of deception detection training relative to a no-treatment control group on a measure of judgment accuracy. Thus, any studies whose reports did not allow the reconstruction of a precise statistical test could not be included in the present effort. It should be noted that, in an effort to obtain the most complete meta-analytic database, attempts were made to contact the authors of those studies that provided incomplete or insufficient data for extraction of statistical tests. Hypothesis tests were coded as having a positive direction of effect if the deception detection training rendered an increase in accuracy compared to the no-training control, and as having a negative direction of effect if the deception detection training rendered a decrease in accuracy compared to the no-training control.

It may be useful to delineate the various types of studies that failed to meet these criteria and which were not included in the meta-analytic database. Studies were excluded if the experimental design did not include a separate no-training control group (e.g. Akehurst et al., 2004; Crews et al., 2007). Studies were excluded if deception detection training was one component of a larger training curriculum (e.g. Marett, Biros, & Knode, 2004). Studies were excluded if participants were simply exposed to videotapes of persons lying or telling the truth (e.g. Mann, Vrij, & Bull, 2006), or were simply told to pay attention to how the speakers looked or what the speakers said (e.g. DePaulo, Lassiter, & Stone, 1982), but there was no actual training intervention involving information, practice, and/or feedback.

In addition to the basic statistical information, each hypothesis test was coded for the predictors described earlier: type of training, training content, motivation of the deceiver, trainee expertise, type of lie, and length of training. Because two persons (the author and a graduate student) coded each variable independently and all disagreements were resolved through discussion, we did not calculate formal estimates of reliability (cf. Patall, Cooper, & Robinson, 2008).[2]

According to the selection criteria, a total of 16 studies with 30 separate hypothesis tests of the effects of deception detection training on judgment accuracy were included in this analysis, representing the responses of 2847 participants. Table 2 presents the hypothesis tests included in this meta-analysis, effect sizes (both d and r are presented), and study moderators. As shown in Table 2, the studies included in this meta-analytic database were published relatively recently, from 1984 to 2006. The typical study involved a sample of approximately $N = 95$ participants, with a range of from 29 participants to 390 participants. In the majority of studies, participants were college students. The length of the training interventions ranged from five to 180 minutes, with a mean of approximately 43 minutes.

Table 2. Studies included in the meta-analysis.

Study	*n*	*d*	*r*	Type of training	Training content: Ten.	For.	Sen.	Eng.	Flu	Motivation	Trainee expertise	Type of lie	Length of training
deTurck (1991)	183	0.698	0.331	Comb.	Y	Y	N	Y	Y	Y	N	N	30
deTurck et al. (1997)													
Visual	82	0.387	0.192	Comb.	Y	N	N	Y	N	Y	N	T	30
Vocal	82	–0.155	–0.078	Comb.	N	Y	N	N	Y	Y	N	T	30
Both	82	0.542	0.264	Comb.	Y	N	N	Y	Y	Y	N	T	30
deTurck et al. (1990)	188	0.541	0.261	Comb.	Y	Y	N	Y	Y	Y	N	T	30
deTurck & Miller (1990)	390	0.719	0.339	Comb.	Y	Y	N	Y	Y	Y	N	N	30
Fiedler & Walka (1993)													
Information	48	1.360	0.569	Info.	Y	N	Y	N	Y	N	N	–	–
Information/feedback	48	1.471	0.599	Comb.	Y	N	Y	N	Y	N	N	–	–
Hartwig et al. (2006)	82	0.969	0.439	Comb.	N	N	N	N	N	N	Y	T	180
Kassin & Fong (1999)	40	–0.672	–0.324	Info.	Y	Y	N	N	N	Y	N	T	40
Kohnken (1987)													
Content	40	0.156	0.079	Info.	N	Y	Y	N	N	Y	Y	T	60
Nonverbal	40	–0.156	–0.079	Info.	Y	N	N	Y	N	Y	Y	T	60
Speech	40	–0.219	–0.111	Info.	N	N	N	N	Y	Y	Y	T	60
Landry & Brigham (1992)	58	0.841	0.392	Comb.	N	Y	Y	N	N	N	N	T	45
Levine et al. (2005)													
Study 2	59	–0.028	–0.014	Info.	Y	N	N	N	Y	N	N	N	5
Study 4	106	0.729	0.345	Info.	N	N	N	N	Y	N	N	N	5

Table 2 (*Continued*)

Study	*n*	*d*	*r*	Type of training	Training content: Ten.	For.	Sen.	Eng.	Flu	Motivation	Trainee expertise	Type of lie	Length of training
Porter et al. (2000)													
Feedback	64	0.736	0.349	Feed.	–	–	–	–	–	Y	N	T	–
Feedback/cue	63	0.672	0.322	Comb.	N	N	N	N	N	Y	N	T	15
Santarcangelo et al. (2004)													
Verbal content	56	0.851	0.396	Info.	N	Y	Y	N	N	N	N	T	–
Vocal	50	0.278	0.140	Info.	Y	N	N	Y	N	N	N	T	–
Visual	51	0.306	0.153	Info.	N	Y	N	N	Y	N	N	T	–
Vrij (1994)													
Information	252	0.191	0.095	Info.	N	N	N	Y	N	N	Y	T	–
Information/feedback	252	0.382	0.188	Comb.	N	N	N	Y	N	N	Y	T	–
Vrij & Graham (1997)													
Students	40	0.849	0.398	Info.	N	N	N	N	N	N	N	T	–
Police	29	–0.339	–0.172	Info.	N	N	N	N	N	N	Y	T	–
Zuckerman et al. (1984)													
Eight-after	64	0.804	0.377	Feed.	–	–	–	–	–	N	N	N	–
Four-after	65	0.000	0.000	Feed.	–	–	–	–	–	N	N	N	–
Four-before	109	0.406	0.201	Feed.	–	–	–	–	–	N	N	N	–
Mixed	67	0.811	0.380	Feed.	–	–	–	–	–	N	N	N	–
Zuckerman et al. (1985)	117	0.687	0.327	Feed.	–	–	–	–	–	N	N	N	–

Note. Type of training: Comb., combined; Info., information-only; Feed., feedback-only.
Training content: Ten., tense; For., forthcoming; Sen., sense; Eng., engaged; Flu., fluent.
Y in the Training content, Motivation, and Trainee expertise columns indicates present; N, absent.
Type of lie: T, transgression; N, non-transgression.

Computation and analysis of effect sizes

The techniques presented by Hedges and Olkin (1985) were implemented in this analysis, using the DSTAT meta-analysis software (Johnson, 1989). Briefly, for each hypothesis test, an effect size g is calculated as the difference between the training and control means, divided by the pooled standard deviation. The gs are converted to ds by correcting them for bias (i.e. g's overestimate of the population effect size, which occurs especially for small study samples). In most cases, the effect size was computed directly from means and standard deviations or from reported F or t statistics. To obtain an overall estimate of the effectiveness of deception detection training, the study outcomes were combined by averaging the d values. Homogeneity, or consistency across studies, was examined by calculating a homogeneity statistic, Q, which has an approximate chi-square distribution with $k - 1$ degrees of freedom, where k is the number of effect sizes.

In the absence of homogeneity, variability in effect sizes was then examined by determining the relationship between study characteristics (moderators) and the magnitude of the effect sizes, using both categorical or continuous models. Categorical models, analogous to analysis of variance (ANOVA), test whether effect sizes differ in magnitude within categories defined by study characteristics. These models produce a between-classes effect, Q_b (analogous to a main effect in an analysis of variance) and a test of the homogeneity of the effect sizes within each class, Q_w. Continuous models, analogous to regression models, test whether study characteristics assessed on a continuous scale are related to the magnitude of effect sizes.

Several of the studies in this database contributed more than one effect size, typically because more than one type of training was implemented. In the meta-analysis reported below, each hypothesis test was treated as an independent observation. This assumption of independence is patently false. For example, from Fiedler and Walka (1993), we derived a test of training (information only) vs control and a test of training (information/feedback) vs control. The effect sizes for these two hypothesis tests are dependent in that the control group for each comparison represents a single control group within this study. However, without making this assumption of non-independence, one would be forced to select the 'best' hypothesis test or to pool the results from the reported hypothesis tests into a single test, an alternative that would seem even more arbitrary and capricious than the present assumption of independence. Moreover, in the present context, this approach would preclude the analysis of moderators of training effectiveness, such as type of training.[3]

Results

General effects

In order to examine the overall efficacy of deception detection training, a weighted mean effect size d_+ was calculated by averaging the d values with each weighted by the reciprocal of its variance (cf. Johnson & Eagly, 2000). This analysis yielded an overall positive effect of deception detection training on judgment accuracy, $d_+ = 0.50$. This corresponds to a 'medium' effect size, according to Cohen's (1988) heuristics for small, medium, and large effect sizes. This overall effect is significant, as evidenced by the 95% confidence interval ranging from 0.42 to 0.057 (excluding zero, the value indicating no difference). A significant test of homogeneity indicates that there is considerable variability around the mean effect size ($Q(29) = 90.03$, $p < 0.0001$). Next, we examine whether the relevant study moderators can explain the variability in training effectiveness.

Effect of type of training

Type of training was shown to be a significant moderator of training effectiveness ($Q_b(2) = 13.24$, $p = 0.001$). As shown in Table 3, combined training produced the strongest effects ($d_{i+} = 0.59$), followed by feedback-only training ($d_{i+} = 0.57$), and information-only training ($d_{i+} = 0.28$). Pairwise comparisons indicated that both combined training and feedback-only training differed significantly from information-only training (ps < 0.001 and < 0.05, respectively).

Effect of training content

We examined whether the presence or absence of cues that have been shown to be significant cues to deception predicted training effectiveness. Results indicated that training was significantly more effective when the training intervention incorporated cues reflecting More Tension (specifically more fidgeting) than when it did not ($Q_b(1) = 4.24$, $p < 0.05$). Training was also significantly more effective when it incorporated cues reflecting Less Sense (sp., less logical, more discrepant) ($Q_b(1) = 11.74$, $p < 0.001$) and cues reflecting Less Fluent (sp., more speech disturbances/word repetitions) ($Q_b(1) = 6.24$, $p < 0.05$). There was a tendency for training to be more effective if it included cues reflecting Less Forthcoming (sp., shorter speaking duration, fewer details), although this difference was not significant ($Q_b(1) = 1.38$, $p > 0.1$). There was little discernable difference in training effectiveness for the cue Less Engaged (sp., fewer illustrators) ($Q_b(1) = 0.01$, $p > 0.1$).

Effect of motivation of the deceiver

There was no significant difference in training effectiveness for studies that provided special incentives to motivate the deceiver versus those that provided no special incentive ($Q_b(1) = 0.15$, $p > 0.1$).

Effect of trainee expertise

Results indicated that training was significantly more effective for those studies involving naive trainees than for studies involving trainees with expertise in detecting deception ($Q_b(1) = 12.16$, $p < 0.001$).

Effect of type of lie

Results indicated that training was significantly more effective for studies in which deception involved non-transgressions (feelings, opinions) than for studies in which deception involved actual transgressions ($Q_b(1) = 9.86$, $p = 0.002$).

Table 3. Tests of categorical models.

Variable and class	Between-classes effect (Q_b)	k	Mean weighted effect size (d_{i+})	95% CI for d_{i+} Lower	Upper	Within-class homogeneity (Q_{wi})[a]
Type of training	13.24*					
Information-only		13	0.28	0.14	0.42	41.24**
Feedback-only		6	0.57	0.38	0.76	7.68
Combined		11	0.59	0.49	0.70	27.87*
Training content						
More tense	4.24*					
Absent		13	0.39	0.27	0.51	34.15**
Present		11	0.56	0.44	0.68	43.20**
Less forthcoming	1.38					
Absent		15	0.43	0.32	0.55	47.74**
Present		9	0.53	0.41	0.66	32.47**
Less sense	11.74**					
Absent		19	0.44	0.34	0.52	59.13**
Present		5	0.92	0.65	1.18	10.71*
Less engaged	0.01					
Absent		15	0.47	0.33	0.61	63.75**
Present		9	0.48	0.38	0.59	17.82*
Less fluent	6.24*					
Absent		13	0.37	0.24	0.49	34.98**
Present		11	0.58	0.47	0.69	40.35**
Motivation of the deceiver	0.15					
No		18	0.51	0.41	0.61	49.91**
Yes		12	0.48	0.37	0.59	39.96**
Trainee expertise	12.16**					
No		23	0.58	0.49	0.66	60.18**
Yes		7	0.27	0.12	0.42	17.06*
Type of lie	9.86*					
Transgression		19	0.36	0.26	0.46	48.04**
Non-transgression		9	0.62	0.50	0.74	15.36

Note. CI, confidence interval.
[a]Significance indicates rejection of the hypothesis of homogeneity.
*$p < 0.05$; **$p < 0.001$.

Effect of length of training

Overall, the length of training was unrelated to training effectiveness ($\beta = 0.002$, $p > 0.1$). Note, however, that the length of training was reported for only half ($k = 15$) of the hypothesis tests in this database.

Discussion

The first goal of this meta-analysis was to establish the efficacy (or lack thereof) of deception detection training. In contrast to the pessimism voiced in most narrative reports, and in keeping with other quantitative evidence (Frank & Feeley, 2003), the results of this research indicate that deception detection training is an effective means of enhancing detection accuracy. In summary, the effect of deception detection training was positive, significant, and of medium magnitude. More broadly, we may compare the overall magnitude of effect of deception detection training on performance ($d_+ = 0.50$) with that obtained from comparable meta-analyses of other training approaches. Recent research indicates that the overall mean effect of team training on performance is $d_+ = 0.59$ (Salas, Nichols, & Driskell, 2007), the overall mean effect of error management training on performance was $d_+ = 0.44$ (Keith & Frese, 2008), and the overall mean effect of organizational training programs on learning criteria was $d_+ = 0.63$ (Arthur et al., 2003). Thus, deception detection training exerts a positive effect on performance comparable to or stronger than these other well-established training interventions.

Why, then, do we observe the skepticism that is often voiced in the literature regarding the effectiveness of deception detection training? First, the overall training effect size is moderate, and there is considerable variability among studies, so it is difficult from a narrative review of the literature to determine the overall effects within this domain. One value of meta-analysis is in abstracting evidence from existing studies, converting them to common metrics, and integrating these results in a rigorously precise manner. The results of this meta-analysis, along with that of Frank and Feeley (2003), indicate that future summaries of the effects of deception detection training should take a more positive form. However, a second reason for the seeming skepticism that people can be trained to more accurately detect deception is that deception detection training has been tested in research settings but not yet proven that it can be effective in real-world environments. This represents the difference between training effectiveness (i.e. does training achieve intended outcomes?) and training transfer (i.e. do the results of training transfer to the real world?). The results of the current study can address the first question, but not the second.

A second goal of this meta-analysis was to examine the extent to which the effectiveness of deception detection training varies as a function of relevant moderators. The results indicate that the effectiveness of deception detection was moderated by the type of training implemented. Combined training, a training intervention that incorporated information, practice, and feedback, produced the largest training effect, followed by training that incorporated feedback only, and training that incorporated information only. Moreover, the strongest type of training, combined deception detection training, rendered an effect size twice that of the weakest type of training, information-only. There are several conclusions that can be drawn from this analysis. First, as discussed earlier, a well-designed training program incorporates several instructional events, including the provision of information to be learned, active practice applying this information, and feedback regarding the correctness of performance. A comprehensive deception detection training intervention would thus incorporate information on cues to deception, practice in identifying these cues, and feedback on accuracy of performance. Although this type of approach is consistent with what we know regarding good training design (see Tannenbaum & Yukl, 1992), and in the current analysis produced the strongest training effect, it is interesting to note that almost half of the training interventions in this database provided information only. Second, although training interventions that incorporated feedback-only rendered an effect similar to that of combined training, the feedback-only results were based on a sample size of only six studies. Third, although combined training was clearly superior to information-only training, all three types of deception detection training were shown to be effective compared to no training.

The results also indicate that the effectiveness of deception detection training was moderated by training content. Clearly, if training is designed to draw attention to behavioral patterns reflecting deception, then it is important that those cues trained are in fact reliable indicators of deception. We selected five categories of cues from DePaulo et al. (2003) that were shown to be significant predictors of deception (yielding a mean summary *d* of 0.25). Overall, the results indicate that deception detection training interventions that incorporated these cues were more effective than interventions that did not. Significant differences were observed for training interventions that incorporated cues related to More Tension (fidgeting), Less Sense (logical, discrepant), and Less Fluent (speech disturbances–word phrase repetitions). Although researchers have argued that training effectiveness is related to the diagnostic value of the cues presented in training, this is the first empirical evidence supporting this claim. It is worthwhile to note that whereas the DePaulo et al. (2003) meta-analysis is the most comprehensive analysis that exists on cues to deception, it does have limitations. For example, almost all of the studies that are included in the DePaulo et al. meta-analytic database represent laboratory studies of lies of little consequence (i.e. 'white lies'). It is possible that the cues to deception identified in DePaulo et al. may differ from cues to deception in high-stakes situations. Nevertheless, the DePaulo et al. results provide the best available estimate of reliable cues to deception.

Contrary to expectations, the results indicate that the effectiveness of deception detection training was not moderated by the motivation of the deceiver. DePaulo et al. (2003) found that cues to deception were stronger when deceivers were motivated to succeed than when there was no special motivation, and thus we expected that training would be more difficult under the conditions in which liars were unmotivated and cues less evident. The present results do not support this argument and indicate that the positive effect of deception detection training is similar for both motivated and unmotivated deceivers. However, it should be noted that the distinction between motivation and no-motivation in typical laboratory studies of deception detection training is relatively modest. That is, in this database, some studies did not provide any special motivation to deceivers, whereas those that provided special motivation may have done this by offering a prize or small cash bonus if deception was successful. Clearly, this may not reflect the level of motivation or consequences inherent in many real-world settings of interest.

Trainee expertise was found to be a significant moderator of training effectiveness, with the results indicating that training was more effective for naive trainees than for trainees with expertise in deception detection. This is not altogether unexpected. Research suggests that both naive persons and experienced professionals harbor many misconceptions regarding cues to deception (Vrij, 2000). That is, cues that people believe to be most stereotypically indicative of deception are often those that are unrelated to deception. For example, research indicates that gaze aversion is one of the most widely held deception stereotypes (Global Deception Research Team, 2006), yet DePaulo et al.'s (2003) results indicate little association between gaze aversion and deception (mean $d = 0.03$). Moreover, some have noted that the commonly used commercial training materials used to train professionals foster these same stereotypes, and that the information provided in professional training programs is questionable and may even be counterproductive (Mann et al., 2004). To the extent that these stereotypes are more deeply entrenched in professionals through inappropriate training and acculturation than in laypersons, the views of professionals may be more difficult to change. In other words, it may be easier to train those with vague misconceptions than those with more ingrained misconceptions, although this is clearly speculative at this point.

Results also indicated that the effectiveness of deception detection training was moderated by the type of lie told by the deceiver. Although deception detection training was effective for both transgressions (lies about one's actions) and non-transgressions (lies about ones' feelings or opinions), training was shown to be more effective when lies were about non-transgressions than about transgressions.

Given these patterns, it may be informative to speculate on how to account for the positive effects of deception detection training. First, drawing on existing research on cues to deception, although DePaulo et al. (2003) concluded that the looks and sounds of deception are faint, their results identified some behavioral cues that are reliable indicators of deception. When training incorporates these cues, especially in a comprehensive training intervention incorporating information, practice, and feedback, we may expect positive and significant gains in deception detection accuracy. How does deception detection training work? As noted earlier, research shows that people tend to hold misconceptions regarding the behavioral cues that reflect deception, and these misconceptions are widespread and even pancultural (Global Deception Research Team, 2006). A primary goal of training is to shift the focus from these global stereotypes or heuristics to more reliable indicators of deception (see Fiedler & Walka, 1993; deTurck et al., 1997). Comprehensive training interventions that provide information on relevant cues, practice in detecting deception, and feedback to correct errors and identify correct responses can serve to counter inaccurate stereotypes and foster attention to relevant behaviors.

It is likely that deception detection training serves a secondary purpose as well. The results of this research have shown that deception detection training was most effective when the training intervention incorporates information, practice and feedback and when the training content incorporates reliable cues to deception, but the results also indicated that training was effective, although less so, when these conditions do not hold. Indeed, Levine et al. (2005) found that, under certain conditions, bogus training that did not include any cues of diagnostic utility led to improved accuracy relative to a no-training control group. Similarly, DePaulo et al. (1982) found that simply telling study participants to pay attention to tone of voice led to increased detection accuracy. It is likely that these results reflect what Driskell et al. (2008) have called the *indoctrination* component of training – that is, one preliminary goal of training is to focus trainee attention on the task and increase motivation. As Levine et al. (2005) noted, training alone, regardless of content, may lead persons to process information more critically. Therefore, at a minimum, training can call attention to the task of accurately attending to nonverbal information. At best, training can inform, guide, and foster skill development to enhance detection accuracy.

Limitations and implications for future application

There are limitations to our analysis, and to the implications that can be drawn from it. A primary limitation is inherent within the deception detection literature itself. Academic research on deception has primarily focused on everyday lies, or lies of low to moderate consequence. As noted in the introduction to this paper, a recent emphasis has been placed on detecting deception in applied settings such as security checkpoints, a situation in which there is much higher risk and higher consequence for the liar. The question is to what extent does existing research apply to this type of real-world environment? This is a difficult question to answer. Researchers have attempted to develop more realistic laboratory scenarios, including mock theft experiments, in which the deceptive acts are more substantial and the consequences greater than in typical research studies. This problem of ecological validity is not unique to the current study – Howell (1998) has noted that the general research strategy of conducting applied research using abstractions of real-world tasks can serve as a bridge between laboratory experimentation and application in the field. Nevertheless, there is a substantial difference between a laboratory setting and one in which someone is attempting to smuggle contraband past a security checkpoint. The studies that comprise the current meta-analysis represent 'best efforts' to examine deception detection training in realistic laboratory environments. The available evidence indicates that deception detection training is effective. Further research is needed to test the extent to which these results transfer to specific real-world settings of interest.

Although the focus of this research was on deception detection *training*, there is much we do not know regarding deception. For example, there remains considerable controversy in the broader deception literature on what behavioral cues are reliable indicators of deception. For example, whereas, DePaulo et al. (2003) found that the cue *fidgeting* was a significant predictor of deception of modest magnitude ($d = 0.16$), others claim that fidgeting is unrelated to deception (Mann et al., 2004). Moreover, as noted previously, cues to deception in laboratory studies of moderate consequence may differ from those in high-stakes settings. The current study is not able to resolve these issues. However, further research on cues to deception, and on conditions under which certain cues may predict deception, should lead to the development of more reliable training content.

Although notable advances have been made in recent years in the science of training (see Salas & Cannon-Bowers, 2001), most of the training studies in this database are quite basic or rudimentary in terms of training design, strategies, and delivery. In a typical study, trainees may first read a list of cues to deception, practice viewing videotapes and making deception judgments, and are then tested for accuracy in discerning deception. Past deception detection training research involved very little application of advanced instructional strategies or technologies. For example, Salas and Cannon-Bowers (2001) have called attention to the pre-training environment and noted that trainee's previous experiences with training can impact learning outcomes. This may be especially important when training those who have had previous professional deception detection training that may have provided information counter to that provided in a more 'evidence-based' training intervention. Thus, it may be useful to focus training efforts on overcoming previous deception stereotypes as well as on attending to more reliable cues to deception. Others have suggested approaches to optimize learning via event-based simulations (Driskell et al., 2008; Salas & Cannon-Bowers, 2001) and other advanced training technologies. The fact that deception detection training is useful, even given such a basic manner in which it has been implemented to date, leads to optimism regarding what can be achieved in the future.

Conclusions

An integrative meta-analysis can serve several purposes. First, it can provide a precise summary of overall effects within a research domain. The current results indicate that the effect of deception detection training was positive, significant, and of medium magnitude. Second, meta-analysis can reveal factors that increase or decrease training effectiveness. The results indicate that the effectiveness of deception detection training was moderated by (a) the type of training implemented – training was most effective when it incorporated information, practice, and feedback, (b) training content – training was most effective when it incorporated reliable indicators of deception, (c) trainee expertise – training was more effective for naive trainees than for experienced trainees, and (d) the type of lie told – training was more effective when deception involved feelings and opinions than when deception involved transgressions. Third, meta-analysis can often identify what we *do not* know regarding a research domain. In this case, we are able to address questions of training effectiveness (i.e. does training achieve intended outcomes?), but existing research does not address questions of training transfer (i.e. do the results of training transfer to real-world settings of interest?). Finally, meta-analysis can serve to direct attention to potentially profitable areas of research. The current study suggests the value of conducting continued research on deception detection training.

Notes

1. Note that the correspondence between the cues described in DePaulo et al. (2003) and the cues described in the studies in this database was not exact. For example, the DePaulo et al. analysis contains the cue 'fidgeting', as well as the cues, 'self-fidgeting' and 'facial fidgeting.' For our purposes, if a training study included a cue related to fidgeting (e.g. self-adaptors, grooming gestures), we simply coded it as fidgeting.
2. For most of the moderators examined (e.g. motivation of the deceiver, trainee expertise, length of training), items were directly coded from the written reports. Other moderators, such as the type of training implemented, required some judgment on the basis of the two raters. For these moderators, initial disagreements occurred on a small number or cases (less than approximately 10%) and these disagreements were resolved through discussion to achieve complete agreement.
3. The assumption that each of the 30 hypothesis tests represented an independent observation is false. However, it can be seen that such an assumption does not seem to render an inflated summary of this research domain. Consider the results of a supplementary meta-analysis of wholly independent effects, in which multiple hypothesis tests obtained from a single study were combined into a single test. This solution precludes

the examination of the effects of moderators such as type of training or training content, but it does eliminate the problem of non-independence. This produced 16 distinct, independent hypothesis tests, one from each includable study. The results of this supplemental meta-analysis revealed the same pattern as the general effects reported, with an overall positive effect of deception detection training on judgment accuracy, $d_+ = 0.54$ and 95% confidence interval ranging from 0.45 to 0.64.

Acknowledgement

I acknowledge the contributions of Susan E. Brandon, Department of Defense, and Ruth Willis, Naval Research Laboratory.

References

Akehurst, L., Bull, R., Vrij, A., & Kohnken, G. (2004). The effects of training professional groups and lay persons to use criteria-based content analysis to detect deception. *Applied Cognitive Psychology*, *18*, 877–891.

Arthur, W.A., Bennett, W., Edens, P.S., & Bell, S.T. (2003). Effectiveness of training in organizations: A meta-analysis of design and evaluation features. *Journal of Applied Psychology*, *88*, 234–245.

Bond, C.F., & DePaulo, B.M. (2006). Accuracy of deception judgments. *Personality and Social Psychology Review*, *10*, 214–234.

Bull, R. (2004). Training to detect deception from behavioural cues: Attempts and problems. In P.A. Granhag & L.A. Stromwall (Eds.), *The detection of deception in forensic contexts* (pp. 251–268). London: Cambridge University Press.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed). Hillsdale, NJ: Erlbaum.

Crews, J.M., Cao, J., Lin, M., Nunamaker, J.F., & Burgoon, J.K. (2007). A comparison of instructor-led vs. web-based training for detecting deception. *Journal of STEM Education*, *8*, 31–40.

DePaulo, B.M. (1994). Spotting lies: Can humans do better? *Current Directions in Psychological Science*, *3*, 83–86.

DePaulo, B.M., Lassiter, G.D., & Stone, J.I. (1982). Attentional determinants of success at detecting deception and truth. *Personality and Social Psychology Bulletin*, *8*, 273–279.

DePaulo, B.M., Lindsay, J.J., Malone, B.E., Muhlenbruck, L., Charlton, K., & Cooper, H. (2003). Cues to deception. *Psychological Bulletin*, *129*, 74–118.

deTurck, M.A. (1991). Training observers to detect spontaneous deception: Effects of gender. *Communication Reports*, *4*, 81–89.

deTurck, M.A., Feeley, T.H., & Roman, L.A. (1997). Vocal and visual cue training in behavioral lie detection. *Communication Research Reports*, *14*, 249–259.

deTurck, M.A., Harszlak, J.J., Bodhorn, D.J., & Texter, L.A. (1990). The effects of training social perceivers to detect deception from behavioral cues. *Communication Quarterly*, *38*, 189–199.

deTurck, M.A., & Miller, G.R. (1990). Training observers to detect deception: Effects of self-monitoring and rehearsal. *Human Communication Research*, *16*, 603–620.

Driskell, J.E., & Mullen, B. (2005). The efficacy of naps as a fatigue countermeasure: A meta-analytic integration. *Human Factors*, *47*, 360–377.

Driskell, J.E., Salas, E., Johnston, J.H., & Wollert, T.N. (2008). Stress exposure training: An event-based approach. In P.A. Hancock & J.L. Szalma (Eds.), *Performance under stress* (pp. 271–286). London: Ashgate.

Driskell, J.E., Willis, R., & Copper, C. (1992). Effect of overlearning on retention. *Journal of Applied Psychology*, *77*, 615–622.

Fiedler, K., & Walka, I. (1993). Training lie detectors to use nonverbal cues instead of global heuristics. *Human Communication Research*, *20*, 199–223.

Frank, M.G., & Feeley, T.H. (2003). To catch a liar: Challenges for research in lie detection training. *Journal of Applied Communication Research*, *31*, 58–75.

Global Deception Research Team. (2006). A world of lies. *Journal of Cross-Cultural Psychology*, *37*, 60–74.

Hartwig, M., Granhag, P.A., Stromwall, L.A., & Kronkvist, O. (2006). Strategic use of evidence during police interviews: When training to detect deception works. *Law and Human Behavior*, *30*, 603–619.

Hedges, L.V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Orlando, FL: Academic Press.

Howell, W.C. (1998). When applied research works. In J.A. Cannon-Bowers & E. Salas (Eds.), *Making decisions under stress: Implications for individual and team training* (pp. 415–425). Washington, DC: American Psychological Association.

Johnson, B.T. (1989). *DSTAT: Software for the meta-analytic review of research literatures*. Hillsdale, NJ: Erlbaum.

Johnson, B.T., & Eagly, A.H. (2000). Quantitative synthesis of social psychological research. In H.T. Reis & C.M. Judd (Eds.), *Handbook of research methods in social and personality psychology* (pp. 496–528). London: Cambridge University Press.

Kassin, S.M., & Fong, C.T. (1999). I'm innocent!': Effects of training on judgments of truth and deception in the interrogation room. *Law and Human Behavior*, *23*, 499–516.

Keith, N., & Frese, M. (2008). Effectiveness of error management training: A meta-analysis. *Journal of Applied Psychology*, *93*, 59–69.

Kohnken, G. (1987). Training police officers to detect deceptive eyewitness statements: Does it work? *Social Behavior*, *2*, 1–17.

Landry, K.L., & Brigham, J.C. (1992). The effect of training in criteria-based content analysis on the ability to detect deception in adults. *Law and Human Behavior*, *16*, 663–676.

Levine, T.R., Feeley, T.H., McCornack, S.A., Hughes, M., & Harms, C.M. (2005). Testing the effects of nonverbal behavior training on accuracy in deception detection with the inclusion of a bogus training control group. *Western Journal of Communication*, *69*, 203–217.

Mann, S., Vrij, A., & Bull, R. (2004). Detecting true lies: Police officers' ability to detect suspects' lies. *Journal of Applied Psychology*, *89*, 137–149.

Mann, S., Vrij, A., & Bull, R. (2006). Looking through the eyes of an accurate lie detector. *Journal of Credibility Assessment and Witness Psychology*, *7*, 1–16.

Marett, K., Biros, D.P., & Knode, M.L. (2004). Self-efficacy, training effectiveness, and deception detection: A longitudinal study of lie detection training. *Lecture Notes in Computer Science*, *3073*, 187–200.

Meissner, C.A., & Kassin, S. (2002). He's guilty!': Investigator bias in judgments of truth and deception. *Law and Human Behavior*, *26*, 469–480.

Mullen, B. (1989). *Advanced BASIC meta-analysis*. Hillsdale, NJ: Erlbaum.

Patall, E.A., Cooper, H., & Robinson, J.C. (2008). The effects of choice on intrinsic motivation and related outcomes: A meta-analysis of research findings. *Psychological Bulletin*, *134*, 270–300.

Porter, S., Woodworth, M., & Birt, A.R. (2000). Truth, lies and videotape: An investigation of the ability of federal parole officers to detect deception. *Law and Human Behavior*, *24*, 643–658.

Rosenthal, R., Hiller, J.B., Bornstein, R.F., Berry, D., & Brunell-Neuleib, S. (2001). Meta-analytic methods, the Rorschach, and the MMPI. *Psychological Assessment*, *13*, 449–451.

Salas, E., & Cannon-Bowers, J.A. (2001). The science of training: A decade of progress. *Annual Review of Psychology*, *52*, 471–499.

Salas, E., Nichols, D., & Driskell, J.E. (2007). Testing three team training strategies in intact teams: A meta-analysis. *Small Group Research*, *38*, 471–488.

Santarcangelo, M., Cribbie, R.A., & Hubbard, A. (2004). Improving accuracy in veracity judgment through cue training. *Perceptual and Motor Skills*, *98*, 1039–1048.

Tannenbaum, S.I., & Yukl, G. (1992). Training and development in work organizations. *Annual Review of Psychology*, *43*, 399–441.

Vrij, A. (1994). The impact of information and setting on detection of deception by police detectives. *Journal of Nonverbal Behavior*, *18*, 117–136.

Vrij, A. (2000). *Detecting lies and deceit: The psychology of lying and the implications for professional practice*. New York: Wiley.

Vrij, A., & Graham, S. (1997). Individual differences between liars and the ability to detect lies. *Expert Evidence*, *5*, 144–148.
Zuckerman, M., Koestner, R., & Alton, A.O. (1984). Learning to detect deception. *Journal of Personality and Social Psychology*, *46*, 519–528.
Zuckerman, M., Koestner, R., & Colella, M.J. (1985). Learning to detect deception from three communication channels. *Journal of Nonverbal Behavior*, *9*, 188–194.