APPLIED COGNITIVE PSYCHOLOGY
Appl. Cognit. Psychol. **20**: 421–446 (2006)
Published online 16 March 2006 in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/acp.1190

Paraverbal Indicators of Deception: A Meta-analytic Synthesis

SIEGFRIED LUDWIG SPORER* and BARBARA SCHWANDT

University of Giessen, Giessen, Germany

SUMMARY

This meta-analysis provides a quantitative synthesis of paraverbal indicators of deception as a function of different moderator variables. Of nine different speech behaviours analysed only two were reliably associated with deception in the weighted, and four in the analysis unweighted by sample size. Pitch, response latency and speech errors were positively, message duration negatively related to deception. As most effect sizes were found to be heterogeneous, analyses of moderator variables revealed that many of the observed relationships varied as a function of content, preparation, motivation, sanctioning of the lie, experimental design and operationalization. Of different theoretical approaches reviewed, a working memory model of lie production may best account for the findings. Because of the small effect sizes, and the heterogeneity in findings, practitioners must be cautioned to use such indicators in assessing the truthfulness of reports but nonetheless practical implications for different types of situations are outlined.

INTRODUCTION

Many people believe that they are able to detect a lie when somebody is trying to deceive them. In particular, professionals who have to deal with (potential) lies as part of their job are confident that their skills at detecting lies are fairly good, and that they are able to catch a liar when necessary. However, a study by Ekman and O'Sullivan (1991) showed that police officers, customs inspectors, federal law enforcement officers, federal polygraphers, robbery investigators, judges and psychiatrists are no more skilled at detecting deception than the average observer. In fact, in this and many other studies, people are barely better than chance in detecting deception. Also, confidence in detecting a lie appears to be unrelated to accuracy in detecting deception (DePaulo, Charlton, Cooper, Lindsay, & Muhlenbruck, 1997). Many observers seem to rely on nonverbal cues when trying to detect deception (e.g. Hale & Stiff, 1990; Vrij, Dragt, & Kopelaar, 1992). However, they may either have difficulty to do so effectively (in ongoing interactions), or the cues they use may not be valid (cf. Vrij, 2000).

*Correspondence to: Prof. Siegfried L. Sporer, Department of Psychology and Sports Science, University of Giessen, Otto-Behaghel-Strasse 10 F, 35394 Giessen, Germany. E-mail: sporer@psychol.uni-giessen.de

Contract/grant sponsor: Deutsche Forschungsgemeinschaft; contract/grant number: (Sp 262/3-2).

Cues can be subdivided into three categories. Nonverbal visual cues such as eye contact, head movements, or hand, leg and foot movements can be observed in ongoing interactions. Paraverbal cues are vocal cues that accompany speech behaviour such as voice pitch, response latencies and filled and unfilled pauses. Apart from paraverbal cues in the auditory channel, there are also verbal content cues, for example the use of certain word classes (e.g. first person pronouns), the immediacy or logical consistency of a statement, or certain types of details contained in a statement. Of these various cues, the focus here is on paraverbal cues to deception.

Past research on cues to deception has, on the whole, produced inconsistent and contradictory results. Zuckerman, DePaulo, and Rosenthal (1981) conducted a meta-analysis that led to the conclusion that of nine cues in the auditory channel, five were significantly positively associated with deception, namely speech errors, speech hesitations, pitch, negative statements and irrelevant information, the latter two being content cues. A later meta-analysis by Zuckerman and Driver (1985) found that response length was significantly negatively, and speech errors, speech hesitations and pitch significantly positively related to deception. The results of both meta-analyses suggest that some paraverbal behaviours might be helpful in distinguishing lie and truth, but on the whole only small to moderate effect sizes were observed. The effect size for pitch was somewhat higher but pitch changes can only be detected with instrumental aids, not in ongoing face-to-face interactions.

Recently, DePaulo et al. (2003) analysed 158 different cues to deception, including a large variety of nonverbal, paraverbal and verbal cues, as well as certain content aspects of deceptive and truthful messages in a very comprehensive meta-analysis.[1] Of the paraverbal behaviours investigated by DePaulo et al. (2003), only three were significantly associated with deception. Pitch ($d = 0.21$) and vocal tension ($d = 0.26$) were higher, and talking time ($d = -0.35$) was shorter in deceptive statements than in truthful ones.

However, many of the behaviours investigated by DePaulo et al. showed outcomes that varied widely from study to study (as indicated by significant rejections of the null hypothesis in homogeneity tests). An explanation of the contradictory findings obtained between these studies might be that studies differed from one another regarding experimental method, the type of sample or the operationalizations used to measure the nonverbal behaviours of interest. With respect to verbal and paraverbal indicators, it is likely to make a difference whether or not a statement could be planned beforehand. Some researchers had given their participants the opportunity to prepare their deceptive messages, others had not. Also, it is quite difficult to motivate participants to lie in a laboratory setting where the consequences appear minimal compared to real life situations, especially in criminal proceedings and before courts of law. Therefore, the opportunity to prepare an account and motivation may be important moderator variables regarding the ecological validity of deception cues. The content of a message may also be quite important with respect to paraverbal correlates of deception. It could make a difference whether a lie is about the communicator's personal feelings or attitudes, or about a factual event.

In this article, we present a different meta-analysis on a small subset of these behaviours, that is the paraverbal indicators message duration, number of words, speech rate, response latency, unfilled pauses, filled pauses, speech errors, repetitions and pitch.

[1]Unfortunately, we were not aware that DePaulo et al. had embarked in this large scale meta-analysis until all our data collection, analyses and first draft of our write-up were completed. Therefore, we have integrated only some of their findings in our introduction and discussion.

We also focus on practical implications of our findings, particularly in forensic contexts, by contrasting objective indicators of deception with lay persons' beliefs or assumptions about changes in these behaviours (i.e. subjective indicators of deception).

Our meta-analysis considers a larger variety of moderator variables to account for differences in the findings (see Lipsey, 1994) than previous meta-analyses. In particular, we emphasize the importance of preparation and planning (including rehearsal) which may be especially relevant in criminal interrogations. While DePaulo et al. (2003) contrasted only those few studies (with a maximum $k = 6$ studies) which had manipulated preparation directly, we coded all studies with respect to the presumed preparation time available to participants (leading to *k*s up to 23 studies).

In addition to preparation, the variables motivation, the content of the deceptive message, sanctioning (sanctioned vs. unsanctioned lies), degree of interaction between experimenter and participant and type of experimental design and operationalizations used were investigated as potential moderators.

Meta-analytical procedures provide ways to assess the magnitude of effects of moderator variables on behavioural correlates of deception. This means, if there is considerable variability in study outcomes, this may be due to the influence of other variables that moderate the relation between deception and the dependent variables (Mullen, 1989). Zuckerman and Driver (1985) presented effect sizes for behaviours associated with deception separately for different levels of planning, the content of the lie and levels of motivation. Response latency and speech rate were associated differently with deception under different levels of planning. When there was little opportunity for planning, liars took more time before answering and spoke slower than truth-tellers. Given high levels of preparation, the relationship was just the opposite. DePaulo et al. (2003) found comparable reliable differences as a function of planning only for response latencies based on the few studies that had manipulated this variable experimentally. In contrast, in the present meta-analysis *all* studies were coded regarding the (presumed) amount of preparation available to participants.

Zuckerman and Driver (1985) reported shorter responses and a slower speech rate as well as a higher pitch of liars under high levels of motivation. However, there were no reliable relationships under low levels of motivation for these variables. DePaulo et al. (2003) confirmed these findings for pitch. They also observed that the number of filled pauses ('ah', 'um', etc.) and of 'non-ah disturbances' (other speech errors such as grammatical errors, false starts, etc.) associated with deception differed as a function of motivation. Under low motivation these two behaviours were nonsignificantly positively associated with deception. In contrast, under high (instrumental) motivation, these types of errors were nonsignificantly negatively related.

DEFINING PARAVERBAL INDICATORS OF DECEPTION

In a typical paradigm for studying deceptive communication, two groups of communicators are formed (by random assignment). One group of individuals is instructed to answer truthfully, the second group receives instructions to falsify their answers. In order to reduce between-subjects variability, sometimes a within-subjects design is chosen, i.e. participants of the same group are observed repeatedly. Behaviours are measured in both groups and the mean values for truthful and deceptive messages are compared.

We carefully scrutinized the operational definitions given in available studies. Unlike DePaulo et al. (2003), we reduced the number of previously examined behaviours by classifying them into groups of equivalent measures, for example by combining different types of pauses. For analysing paraverbal behaviours, audio recordings (and standardized transcripts thereof) are necessary. The definitions of most behaviours analysed were quite straightforward. For example most studies included provided data for response latency. In almost all of these studies, this variable was operationalized identically as the time between the end of the interviewer's question and the beginning of the participant's answer (e.g. Greene, O'Hair, Cody, & Yen, 1985).

One can distinguish single measures such as message duration (talking time) or number of words (frequency count) from composite measures such as speech rate (number of words divided by message duration). Filled pauses denote the frequency of speech disturbances such as uh, er, um, ah, etc., while unfilled pauses were measured as duration or frequency of silent breaks in speech. Repetitions refer to the frequency of words or phrases that are duplicated in a series and are not interrupted by a pause or a speech error, or by a question from the interviewer. Speech errors refer to the frequency of grammatical errors (switching plural/singular, omissions of verbs, omissions of contractions), stuttering, sentence fragments/false starts, etc. Frequency measures are usually reported as relative frequencies (per minute or text length). Pitch was usually measured as the fundamental frequency of the voice analysed by electronical devices.

The purpose of the present meta-analysis was not only to provide an integrative summary of paraverbal correlates of deception as a function of a series of moderator variables but also to test theory-driven predictions. In the following, we outline several approaches which have been suggested to account for behavioural differences when lying in contrast to telling the truth.

PREDICTIONS FROM DIFFERENT THEORETICAL APPROACHES

Many researchers have drawn upon four distinct approaches which may account for paraverbal (and nonverbal in general) differences when lying in contrast to telling the truth (e.g. Koehnken, 1990; Vrij, 2000; Zuckerman et al., 1981; Zuckerman & Driver, 1985). Either approach has its own specific predictions regarding paralinguistic behaviours as shown in Table 1 but the four factors may operate in concert, i.e. these behaviours are considered to be multi-determined (Zuckerman et al., 1981). The first three approaches are only described briefly as they have been described in detail by previous authors. However, we offer a different theoretical rationale for the 'cognitive load approach' by couching it into a working memory model of lie production which we consider particularly useful for the construction of complex lies. Note that these approaches make partially contradictory predictions (see Table 1). Table 1 also includes assumptions about these paraverbal cues that are believed to be associated with deception. These beliefs have been shown to be held both by lay people and professionals alike (Akehurst, Koehnken, Vrij, & Bull, 1996; Breuer, Sporer, & Reinhard, 2005; Koehnken, 1990; Stromwall, Granhag, & Hartwig, 2004; Vrij, 2000).

The arousal, affective, and attempted control approach

There is the general notion that whenever faced with an unusual, threatening, or complex situation individuals experience a greater degree of arousal (Kahneman, 1973). The physiological responses associated with nonspecific arousal are difficult to control and

Table 1. Predictions regarding changes in nonverbal and paraverbal behaviours when lying derived from different theoretical approaches and lay assumptions

Behaviour	Arousal	Emotion/ affect		Attempted control	Cognitive load/ Working memory	Lay assumptions
		Fear	Guilt			
Message duration	?	<	?	> | <	<	?
Number of words	?	<	?	> | <	<	?
Speech rate	>	>	<	> | <	<	>
Filled pauses	>	>	?	<	>	>
Unfilled pauses	?	>	?	<	>	>
Pitch	>	>	? | <	?	?	>
Repetitions	? | >	>	?	?	>	?
Response latency	? | <	?	>	<	>	>
Speech errors	>	?	?	? | <	>	>

Note: > = increase with deception; < = decrease with deception;
? = no prediction available or prediction not clear;
| = rival predictions depending on certain aspects of an approach.

therefore may provide relatively consistent cues to deception (Zuckerman et al., 1981). Hence, we would expect to see an increase in frequency of signs of autonomic activity, such as pitch, and in filled pauses and speech errors (Zuckerman & Driver, 1985).

Arousal is frequently associated with specific emotions,[2] some of which are likely to be experienced by liars, such as fear of punishment or guilt. Ekman (1988) argued that the fear of being caught is highest when the liar expects severe punishment in the case of being caught, when the liar has not practiced the lie before, and when the interrogator is suspicious. A person may feel guilty about lying which should lead to slower speech. Note that depending on the emotion involved, increases or decreases of some of the behaviours would be predicted (for example for speech rate see Table 1). Although predictions are not always clear cut, we would expect liars who fear being caught to construct shorter messages (duration and length) and to speak more slowly to avoid contradictions. Whether or not fear or guilt will also be accompanied by more speech errors or any other paraverbal indicators cannot directly be predicted.

The attempted control approach refers to the supposition that liars should try to control the different communication channels in order to create a credible demeanour (Ekman, 1981; Ekman & Friesen, 1969; Zuckerman et al., 1981). This notion presupposes that the sender of the deceptive message should have an idea of the behavioural displays presumably raising scepticism about his or her own credibility in the receiver (see also Koehnken, 1990). Specifically, individuals often associate deception solely with increasing (not decreasing) intensity or frequency of certain behaviours (Akehurst et al., 1996; Breuer et al., 2005; Koehnken, 1990; see Table 1) and therefore will try to avoid these behaviours when lying. To the extent that liars believe that answering shortly, speaking slowly and responding slowly will be interpreted as signs of lying, they may try to give more elaborate answers, speak faster, and respond more quickly. Of course, the latter will largely depend on the preparation, planning and rehearsal possible before a given deception situation. On the other hand, if they believe that too detailed answers may

[2]Although we are aware that emotion researchers distinguish between emotion and affect, we use these terms interchangeably as many deception researchers have done (e.g. Zuckerman et al., 1981, refer to this as the affective approach, while citing Ekman's work who refers to specific emotions and feelings).

give them away, liars may show the opposite behaviours. As the tone of voice appears difficult to control (e.g. Zuckerman, Amidon, Bishop, & Pomerantz, 1982), an increased pitch may give liars away. Liars will attempt to avoid pauses and attempt to respond quickly to a question but whether or not they are able to control speech errors as well is unclear. In the following, we present a revision of the fourth, the cognitive load approach in the form of a working memory model of lie production.

The cognitive approach/working memory model

Encoding a deceptive message requires a greater cognitive effort than telling the truth because of higher processing capacity demands (Koehnken, 1990; Miller & Stiff, 1993; Vrij, 2000; Zuckerman et al., 1981), particularly when the lie involves the report about a complex event (Sporer & Zander, 2001; Sporer, Krompass, & Breuer, 2005). For instance, the liar must pay attention not to contradict former statements within the same or a previous interview, or the listener's knowledge (Cody, Marston, & Foster, 1984). Furthermore, the message must seem plausible, which can be achieved, for instance, by mentioning some additional details (Cody et al., 1984). These greater cognitive demands of assembling untruthful messages while talking should result in a change in intensity or frequency of certain behaviours. The extent of these behavioural differences should depend upon the time required to retrieve information from memory (Cody et al., 1984). There is also some evidence at a behavioural level as liars appear to be 'thinking hard' which may actually give them away (Vrij, Edward, & Bull, 2001).

However, the cognitive approach should be more closely linked to contemporary theories of memory and cognition. Originally, the cognitive approach was linked to information processing models (e.g. Atkinson & Shiffrin, 1968) and work on psycholinguistics (e.g. Goldman-Eisler, 1968), which suggested that verbal statements of higher complexity are likely to be associated with longer response latencies and more frequent pauses (Zuckerman et al., 1981; Vrij, 2000). While other researchers have focused more on task demands ('cognitive load'; e.g. Vrij, 2000), our revision is primarily concerned with the cognitive and memorial processes involved in constructing a (complex) lie. A cognitive theory of lie production needs to take current models of working memory (e.g. Baddeley, 2000a, 2000b) and research on autobiographical memory into account (e.g. Anderson & Conway, 1997; Brewer, 1986, 1996; Conway, 1990; Tulving, 1989; Wheeler, Stuss, & Tulving, 1997). When retrieving a (truly experienced) autobiographical memory Baddeley's (2000a) episodic buffer forms the crucial interface between memory and conscious awareness. Thus, it can be related to research on autobiographical memory which postulates retrieval through the process of conscious recollection (Tulving, 1989; Wheeler et al., 1997).

Telling the truth about an experience involves not only retrieval but also reconstruction of an episodic, or autobiographical, memory. This should require some effort at reconstruction, particularly when the to-be-recalled event is complex. In contrast, when constructing a (complex) lie, a person needs to either freely invent (using his or her imagination) and/or to construct a coherent story based on script knowledge (Schank & Abelson, 1977) about comparable situations and events. Thus, a person truthfully reporting an event can draw readily on existing memory structures which should allow for a more elaborate, longer story, faster speech and quicker responses to questions. Skilled liars may achieve some of this as well if the event in question is well scripted, or if they can draw on related experiences. However, when no ready-made script or event-schema (Fiske, 1991) is available, the working memory capacity is more taxed, which would lead to less capacity available for speech production, which in turn should lead to more pauses and speech errors. Finally, recent neuropsychological evidence about the functioning of the frontal cortex suggests that retrieval and reconstruction of complex autobiographical events as well as the monitoring of output provided is likely to be a function of the central executive (Wheeler et al., 1997) which also plays a central role in working memory (Baddeley, 2000a, 2000b). When lies are constructed, skilled liars do not invent complex experiences from scratch but draw upon bits and parcels of past experiences. Just how such information from various sources is integrated can be enlightened by ongoing research on working memory (Baddeley, 2000a; Miyake & Shah, 1999).

The arousal, emotional, control and working memory approaches are considered as complementary rather than rival theoretical approaches (despite us favouring the latter). Nonetheless, they provide a basis for deriving theoretical predictions which we tested in our meta-analysis. For some behaviours the different approaches make similar predictions, in particular the arousal and affective approaches on the one hand, and the attempted control and cognitive load/working memory model on the other hand (see Table 1). But for some of the behaviours investigated here, no specific predictions can be derived from one or the other approach. For other behaviours, the approaches make contrary predictions, which can be tested against each other.

Interestingly, most of the predictions advanced by DePaulo et al.'s (2003) self-presentational approach converge with the ones we have derived from a working memory model of lie production. Whereas DePaulo et al.'s approach is particularly apt for analysing deception in everyday life situations and in interpersonal relationships, the present focus is more on applications of deception research in applied situations, particularly in criminal proceedings and before courts of law. This is also the reason why we investigated a series of moderator variables, which we considered especially important in applied settings.

The present study and predictions

The purpose of the present meta-analysis is (1) to obtain combined effect sizes, along with significance levels, associated with lying for each of the paraverbal variables, (2) to test the predictions of the theoretical approaches to account for differences in paraverbal behaviours when lying versus truth telling, and (3) to examine the impact of practically important moderator variables on these paraverbal behaviours in deception.

Presumably, the most important moderator variable with respect to practical applications is the very content a person is lying about. It should make a big difference whether or not a communicator is lying about his or her feelings and attitudes or about facts of an event. Differences in message length, speech rate and response latencies as a function of deception should more likely emerge regarding complex factual events (as predicted from the working memory model) while arousal and emotion-related differences such as speech errors would be expected more when lying about attitudes and feelings.

The most clear predictions regarding preparation, planning and rehearsal follow from the working memory model. Compared to short preparation conditions, deception under long preparation conditions should be associated with longer message durations, a greater number of words, higher speech rates, but fewer speech errors and shorter response latencies. This hypothesis underlies the assumption that planning of lies should burden working memory capacity less during speech production. Therefore, poorly prepared liars should have less capacity available than well-prepared deceivers, resulting in more obvious changes in the frequency or intensity of certain behaviours.

With respect to the moderator variable motivation we expected that, compared to low motivating conditions, deception under highly motivating conditions should be even more strongly associated with a higher pitch and more speech errors due to arousal as these behaviours are presumed to be difficult to control (the motivational impairment effect; cf. DePaulo & Kirkendol, 1989). From the attempted control approach one would expect that highly motivated liars will try harder to control their behaviour, and therefore speak faster and respond more quickly. In contrast, from the working memory model we would predict that highly motivated liars will be especially careful to scrutinize their answers before responding to avoid contradictions, which should lead to slower response latencies. Lies not sanctioned by the experimenter should result in similar effects as conditions of high motivation (cf. Miller & Stiff, 1993).

Finally, we included several methodological variables as potential moderators, viz. experimental design (between vs. within), type of operationalization (duration vs. frequency), and whether or not participants interacted with an interviewer (or were simply recorded). We expected within-subjects designs to be more sensitive to changes in paraverbal behaviours. Since Buller and Burgoon (1996) have argued that the interaction between sender and receiver is crucial for understanding interpersonal deception we categorized studies by the amount of interaction involved. Differences between liars and truth-tellers should be larger when interactions are two-sided.

Method

Retrieval of studies

First, an on-line search with the Social Sciences Citation Index (SSCI) for the period 1981–2000 was conducted using keywords such as ‘deception’, ‘deceptive’, ‘lie’ and the like. Due to homographs (‘lie’, ‘lies’ etc.) with multiple meanings this search led to a very large database (over 5000 entries), which was carefully sifted on the basis of titles and abstracts. Reference sections of many of these articles were searched, and potentially relevant earlier studies were located. Altogether, 41 studies provided sufficient statistical data to be included in this meta-analytic synthesis. In some of these studies, results of different experimental conditions were considered separately when these conditions differed from one another with respect to the moderator variables.

Coding procedures

Two coders completed all coding necessary for this analysis. Codings included statistical information on mean values, F, t, or r values, sample size, degrees of freedom, significance levels and direction of effect and category assignment of moderator variables. Whenever possible, we attempted to obtain precise estimates of effect sizes from the available indices. Sometimes, the respective test statistics had to be reconstructed from other values reported (e.g. means, cell sizes, and *MSe* values; see Winer, Brown, & Michels, 1991, p. 426). Effect sizes r were calculated according to Rosenthal (1991, equation 2.17). In some cases test statistics were not reported in the primary studies, whereas significance levels were. In these cases we attempted to reconstruct the test statistics from the tables of common statistics textbooks which would lead to conservative, lower estimates of effect sizes (Rosenthal, 1994). For example for a difference in a sample with $N = 100$ reported as ‘significant’ with *alpha* set to 0.05, the effect sizes $|r| = 0.16$ and $|d| = 0.33$ can be reconstructed, with the sign added according to the direction of the effect indicated. In case no statistical tests were reported but the results were described as not significant, *r*s and *d*s were set to 0.00. This was necessary 41 times of all 121 effect sizes analysed.

The coding of moderator variables involved intricate coding operationalizations that were established in advance (cf. Stock, 1994). For most moderator variables, two categories were established. For motivation and preparation, we first created finer distinctions with four and three coding categories, respectively. However, due to small cell sizes, these finer categories had to be abolished by combining them. For all codings of these study descriptors, inter-coder reliabilities were highly satisfactory, with all coefficients indicating either perfect agreement, or being higher than $r = 0.94$. In the few cases of disagreements, the two coders discussed their decisions until reaching agreement.

Results

We first present a simple vote-counting procedure of the significance of study outcomes and inspect the funnel plots to check for publication biases. Then we present the overall unweighted and weighted analyses of mean effect sizes across studies. Finally, we examine in more detail the role of the different moderator variables.

Vote-counting and funnel plots

For 88 of a total of 119 tests no significant differences between lying and truth telling were obtained. Only in 17 studies we observed significant increases, and only in 14 studies significant decreases in paraverbal correlates of deceptive statements. We also drew funnel plots (sample sizes plotted against Fisher's *Z* values of the respective effect sizes) as an eye-balling method for the homogeneity of results across studies (Light & Pillemer, 1984; Mullen, 1989). Only for two variables (unfilled pauses and pitch), a funnel-like shape appeared, with null or small results based on small sample sizes missing. Although these two plots at first glance could be described as a funnel, they appear more to take the shape of an inverted Y, as there were no publications of small study outcomes based on small sample sizes available. Thus, a null results publication bias was probably not operating.

Unweighted analyses

Table 2 displays the number of studies and total sample sizes as well as Fisher's *Z*, unweighted mean *r*, observed variance, fail-safe numbers, 95% confidence intervals and Cohen's *d* in the unweighted analyses. Four paralinguistic variables were associated with deception. Message duration was shorter when participants were lying ($r = -0.057$, $p = 0.018$), pitch was higher during deception ($r = 0.132$, $p = 0.013$), response latency was longer before giving deceptive answers ($r = 0.088$, $p = 0.004$), and speech errors occurred more often during deception ($r = 0.059$, $p = 0.035$).

Weighted analyses

Table 3 shows the results of the weighted analyses using both the Hedges and Olkin and Schmidt-Hunter method. Only two paralinguistic behaviours were associated with deception, pitch ($r = 0.101$, $p = 0.044$), and response latency ($r = 0.106$, $p = 0.001$). At the same time, pitch and response latency effect sizes were highly heterogeneous ($Q = 18.26$, $p = 0.006$; $\chi^2 = 18.15$, $p = 0.006$, and $Q = 108.82$, $p < 0.001$; $\chi^2 = 97.24$, $p < 0.001$, respectively). Heterogeneity statistics for message duration and filled pauses were significant as well ($Q = 75.26$, $p < 0.001$; $\chi^2 = 69.28$, $p < 0.001$ and $Q = 50.89$,

Table 2. Integration of unweighted effect sizes for nonverbal indicators of deception audible in speech

Dependent variable	*k*	Total *N*	Fisher's *Z*	Unweighted mean *r*	*p*	Observed variable	fsn 0.05	95% CI for *r*	Cohen's *d*
Message duration	23	1378	−.057	−.057	.018	.057	3.02	−.110 to −.002	−.113
Number of words	8	350	.012	.012	.409	.028	[a]	−.096 to .120	.025
Speech rate	16	981	.007	.007	.413	.033	[a]	−.057 to .071	.014
Filled pauses	15	762	.026	.026	.233	.081	[a]	−.047 to .099	.053
Unfilled pauses	9	464	.017	.017	.355	.033	[a]	−.076 to .111	.035
Pitch	7	284	.133	.132	.013	.126	11.46	.012 to .248	.268
Repetitions	4	136	.112	.112	.098	.015	4.94	−.064 to .281	.226
Response latency	18	890	.088	.088	.004	.127	13.66	.020 to .155	.177
Speech errors	19	961	.059	.059	.035	.030	3.29	−.006 to .123	.118

Note: *k* represents the number of tests included. Positive *r* values indicate that an increase in the behaviour was associated with deception. Fail-safe numbers are reported for a critical *p* of .05. [a] Fail-safe number is not meaningful due to lack of significance.

$p < .001$; $\chi^2 = 44.00$, $p < .001$, respectively). Furthermore, the assessment of homogeneity via the residual standard deviation approach yielded the conclusion that the effect sizes of unfilled pauses, speech errors, and speech rate are to be considered heterogeneous as well (residual *SD*s are .057, .078, and .081, respectively). Therefore, all paralinguistic variables were subjected to moderator variable analyses, except for repetitions and number of words.

The interrelated variables message duration, number of words and speech rate were not significantly associated with deception. However, the signs of their mean effect sizes illustrate their interrelation. Speech rate slightly increased during telling lies ($r = .012$), whereas the number of spoken words stayed the same ($r = -.009$). This should go hand in hand with a shorter message duration, and according to the Hedges and Olkin (1985) analysis on weighted effect sizes, it did ($r = -.039$, $p = .074$).

Moderator variables: focused comparisons of effect sizes

In order to test the effects of the moderator variables, the technique of classification (or blocking) was chosen (Mullen, 1989).[3] Hypothesis tests were classified ('blocked') into groups, based on the levels of the predictor under consideration (Mullen, 1989). The same analyses as already described were completed for separate groups, yielding a mean *r* and its associated 95% confidence interval for each behaviour within each category of a predictor. Afterwards, one can compare to what extent study outcomes differ, for instance, between experiments with short or medium preparation time. Whether or not the effect sizes in subgroups differ is tested via a *Z*-test (normal distribution) for the difference between correlations (cf. Zuckerman & Driver, 1985). In addition, it is important to determine whether or not the effect size in a particular subgroup differs significantly from zero and to examine its associated confidence interval.

Content of the lie

Originally, we had classified the studies in those involving lies about feelings (and attitudes), about facts only, and about feelings and facts. To obtain sufficient block sizes, the categories feelings and both feelings and facts were collapsed together. Thus, two

[3]Although more sophisticated methods of analyses are now available to consider multiple moderator variables simultaneously (see Lipsey, 1994; Lipsey & Wilson, 2001), cell sizes were too small to conduct these analyses.

Table 3. Integration of weighted effect sizes for nonverbal indicators of deception audible in speech

Dependent variable			Hedges & Olkin				Schmidt-Hunter							
	k	Total N	Fisher's Z	Mean r	p	Q	Fisher's Z	Mean r	p	Observed variable	χ^2	fsn 0.05	95% CI for r	Cohen's d
Message duration	23	1378	−.039	−.039	.074	75.26**	−.040	−.040	.069	.050	69.28**	−4.62	−.094 to .014	−.080
Number of words	8	350	−.009	−.009	.433	7.19	−.008	−.008	.442	.021	7.26	[a]	−.116 to .100	−.016
Speech rate	16	981	.012	.012	.354	22.10	.012	.012	.358	.023	22.48	[a]	−.053 to .076	.023
Filled pauses	15	762	.040	.040	.135	50.89**	.039	.039	.143	.058	44.00**	−3.38	−.034 to .111	.078
Unfilled pauses	9	464	−.016	−.016	.365	10.29	−.014	−.014	.379	.023	10.49	[a]	−.108 to .079	−.029
Pitch	7	284	.102	.101	.044	18.26**	.104	.103	.041	.063	18.15**	7.48	−.017 to .221	−.179
Repetitions	4	136	.084	.084	.164	1.86	.087	.086	.158	.014	1.97	2.92	−.089 to .257	.174
Response latency	18	890	.106	.106	.001	108.82**	.105	.105	.001	.107	97.24**	19.70	.037 to .171	.211
Speech errors	19	961	.041	.041	.102	25.70	.042	.042	.096	.026	24.93	[a]	−.023 to .107	.084

Note: k represents the number of tests included. Positive r values indicate that an increase in the behaviour was associated with deception.
Fail-safe numbers are reported for a critical p of .05. [a] Fail-safe number is not meaningful due to lack of significance.
$^*p < .05$; $^{**}p < .01$.

Table 4. Mean weighted r effect sizes (Hedges & Olkin, 1985) as a function of content of the lie

Variable	Content	k	N	r	Z	95% CI for r
Message duration	Feelings and facts	11	564	.014	1.62	±.084
	Facts only	12	814	−.075*		±.068
Speech rate	Feelings and facts	6	191	−.094	1.61	±.149
	Facts only	10	790	.036		±.072
Filled pauses	Feelings and facts	5	226	.012	0.49	±.133
	Facts only	10	536	.051		±.086
Unfilled pauses	Feelings and facts	5	272	−.003	0.34	±.123
	Facts only	4	192	−.035		±.148
Pitch	Feelings and facts	5	161	.220**	2.27*	±.148
	Facts only	2	123	−.051		±.178
Response latency	Feelings and facts	5	183	.407**	4.89**	±.102
	Facts only	13	707	.024		±.077
Speech errors	Feelings and facts	6	252	.076	0.64	±.126
	Facts only	13	709	.029		±.076

Note: k represents the number of tests included. Positive r values indicate that an increase in the behaviour was associated with deception. Z indicates the significance of the difference between the two correlations (normal distribution).
$^*p < .05$; $^{**}p < .01$.

categories remained: feelings and facts versus facts only (see Table 4). Lying about facts only was associated with a decrease in message duration ($r = -.075$, $p = .016$), but lying about both feelings and facts was not ($r = .014$, $p = .370$). Lying about both feelings and facts was associated with a highly significant increase in pitch ($r = .220$, $p = .003$) and in response latency ($r = .407$, $p < .001$), while lying about facts only showed no such relationships ($r = -.051$, $p = .286$, and $r = .024$, $p = .262$, respectively).

Preparation

None of the studies involved very long preparation times so we compared only behaviours after short (that is minimum preparation) and medium preparation (see Table 5). Lying participants, compared to truth-tellers, tended to talk for a shorter time when they had longer preparation time ($r = -.068$, $p = .025$) while there was no difference with short preparation time ($r = .006$, $p = .445$). This is astonishing because according to the control approach one would expect well-prepared participants to extend their answers in order to appear honest. Although the r's of speech rate were neither significant in the short ($r = .082$, $p = .052$) nor in the medium preparation studies ($r = -.034$, $p = .205$), the difference between both groups was significant ($Z = 1.78$, $p = .038$). Pitch was positively associated with deception only when participants were poorly prepared ($r = .194$, $p = .005$), but not when they had medium preparation time ($r = -.061$, $p = .268$). As one would expect from the working memory model, under conditions of only short preparation time, participants took longer to respond when lying ($r = .190$, $p < .001$) than under medium preparation time ($r = -.044$, $p = .214$).

Rehearsal

We had also attempted to code studies for the opportunity to rehearse their account but only three out of all available hypothesis tests could be assigned to level 2 (explicit rehearsal). In the remainder of studies, no rehearsal was present. Therefore, no analysis of rehearsal effects was possible.

Table 5. Mean weighted *r* effect sizes (Hedges & Olkin, 1985) as a function of level of preparation

Variable	Preparation	*k*	*N*	*r*	*Z*	95% CI for *r*
Message duration	Short	11	541	.006	1.34	±.084
	Medium	12	837	−.068*		±.069
Speech rate	Short	8	394	.082	1.78*	±.096
	Medium	8	587	−.034		±.084
Filled pauses	Short	12	650	.052	0.86	±.079
	Medium	2	72	−.057		±.232
Unfilled pauses	Short	5	288	.034	1.38	±.118
	Medium	4	176	−.099		±.157
Pitch	Short	5	181	.194**	2.06*	±.144
	Medium	2	103	−.061		±.195
Response latency	Short	12	566	.190**	3.38**	±.079
	Medium	6	324	−.044		±.111
Speech errors	Short	13	607	.062	0.73	±.082
	Medium	5	262	.008		±.124

Note: *k* represents the number of tests included. Positive *r* values indicate that an increase in the behaviour was associated with deception. *Z* indicates the significance of the difference between the two correlations (normal distribution).
$^{*}p < .05$; $^{**}p < .01$.

Motivation

The originally planned fine distinctions between four levels of motivation had to be aggregated into two broader categories. Thus, we differentiated between low motivation, operationalized as a small monetary incentive or enlisting the participant as a confederate, and high motivation, that is, stressing the importance of lying for one's professional development, as in Ekman's studies with nursing students (see Table 6). The negative association between deception and message duration was significant for highly motivated liars ($r = -.102$, $p = .023$) but not for low motivated liars ($r = -.016$, $p = .307$). Speech

Table 6. Mean weighted *r* effect sizes (Hedges & Olkin, 1985) as a function of level of motivation

Variable	Motivation	*k*	*N*	*r*	*Z*	95% CI for *r*
Message duration	Low	13	994	−.016	1.43	±.063
	High	10	384	−.102*		±.106
Speech rate	Low	10	698	−.024	1.79*	±.074
	High	6	283	.102*		±.119
Filled pauses	Low	8	432	.036	0.11	±.095
	High	7	330	.044		±.110
Unfilled pauses	Low	3	248	−.073	1.35	±.125
	High	6	216	.053		±.139
Pitch	Low	5	237	.012	3.51**	±.127
	High	2	47	.529**		±.183
Response latency	Low	10	524	.049	2.05*	±.084
	High	8	366	.187**		±.100
Speech errors	Low	7	356	.081	0.94	±.106
	High	12	605	.018		±.082

Note: *k* represents the number of tests included. Positive *r* values indicate that an increase in the behaviour was associated with deception. *Z* indicates the significance of the difference between the two correlations (normal distribution).
$^{*}p < .05$; $^{**}p < .01$.

Table 7. Mean weighted r effect sizes (Hedges & Olkin, 1985) as a function of sanctioning the lie by the experimenter

Variable	Sanction	k	N	r	Z	95% CI for r
Message duration	Unsanctioned	3	144	−.228**	2.42**	±.164
	Sanctioned	20	1234	−.017		±.057
Speech rate	Unsanctioned	2	91	.209*	1.97*	±.194
	Sanctioned	14	890	−.008		±.066
Filled pauses	Unsanctioned	2	99	.325**	3.13**	±.172
	Sanctioned	13	663	−.005		±.077
Unfilled pauses	Unsanctioned	3	136	.088	1.43	±.171
	Sanctioned	6	328	−.059		±.112
Response latency	Unsanctioned	4	159	.244**	1.94*	±.137
	Sanctioned	14	728	.078*		±.074
Speech errors	Unsanctioned	3	144	−.050	1.17	±.172
	Sanctioned	16	817	.057		±.071

Note: k represents the number of tests included. Positive r values indicate that an increase in the behaviour was associated with deception. Z indicates the significance of the difference between the two correlations (normal distribution).
$^*p < .05$; $^{**}p < .01$.

rate was significantly associated with deception under highly motivating conditions ($r = .102$, $p = .043$), but not under low motivation ($r = -.024$, $p = .263$). Similarly, pitch was positively related to deception under high motivation ($r = .529$, $p < .001$), but not under low motivation ($r = .012$, $p = .427$). Compared to the low motivation condition ($r = .049$, $p = .131$), deception under high motivation was associated with longer response latencies ($r = .187$, $p < .001$). For unfilled pauses, differences in the mean effect size were not significant but involved a change in direction of the effect.

Sanctioned vs. unsanctioned lies

In the few studies involving unsanctioned lies, four variables were more strongly associated with deception (see Table 7): message duration ($r = -.228$, $p = .003$), speech rate ($r = .209$, $p = .023$), filled pauses ($r = .325$, $p = .001$) and response latencies ($r = .244$, $p = .001$). None of these behaviours had been significantly associated with deception in the overall analysis. With sanctioned lies, there were no significant associations except for response latencies ($r = .078$, $p = .018$).

Interaction with the experimenter

Interaction was coded at three levels: no interaction between sender and receiver (talking into a video camera), one-sided (communication towards a partner without questions), and two-sided (with questions and answers by receiver and communicator). Most studies involved two-sided communications. For none of the behaviours did the blocking result in significant differences as a function of type of interaction. Therefore, type of interaction does not seem to moderate the association of deception with paraverbal behaviours.

Experimental design and operationalization of dependent variables

Different directions of effect were obtained in between-subjects compared to within-subjects designs (see Table 8). Surprisingly, effects for the majority of paraverbal behaviours were stronger in between-subjects designs. Message duration was significantly

Table 8. Mean weighted r effect sizes (Hedges & Olkin, 1985) as a function of experimental design

Variable	Design	k	N	r	Z	95% CI for r
Message duration	Between	7	388	−.214**	4.12**	±.099
	Within	16	990	.030		±.063
Speech rate	Between	5	322	.157**	3.20**	±.108
	Within	11	659	−.060		±.078
Filled pauses	Between	5	298	.214**	3.94**	±.108
	Within	10	464	−.076		±.094
Unfilled pauses	Between	3	136	.088	1.43	±.171
	Within	6	328	−.059		±.112
Pitch	Between	1	92	.000	1.19	±.205
	Within	6	192	.152*		±.141
Response latency	Between	13	669	.038	3.54**	±.078
	Within	5	221	.304**		±.113
Speech errors	Between	8	471	.007	1.04	±.093
	Within	11	490	.074		±.091

Note: k represents the number of tests included. Positive r values indicate that an increase in the behaviour was associated with deception. Z indicates the significance of the difference between the two correlations (normal distribution).
$^*p < .05$; $^{**}p < .01$.

negatively associated with deception ($r = -.214$, $p < .001$) while both speech rate and filled pauses were significantly positively associated with deception only in between-subjects designs ($r = .157$, $p = .002$, and $r = .214$, $p < .001$, respectively). In contrast, pitch and response latency only showed significant positive associations with deception in within-subjects designs ($r = .152$, $p = .018$ and $r = .304$, $p < .001$, respectively).

Measurement via ratings or as duration versus frequency yielded different directions of effects for several paraverbal behaviours. Filled and unfilled pauses tended to be positively associated with deception when their duration was measured. These analyses indicate that the way variables are operationalized can not only yield effects that differ significantly in magnitude, but also have different directions. However, the comparisons are based on too few studies to allow for strong conclusions.

DISCUSSION

In this meta-analysis, we provided a quantitative summary of paraverbal correlates of deception which can be observed in ongoing interactions between communicator and target. For the nine paraverbal indicators investigated, the overall results were rather disappointing. As already indicated by the preliminary vote-counting analyses, the majority of findings were null-findings, with few indicators showing an increase or decrease when people were lying.

File drawer problem

As already indicated on the basis of the inspection of funnel plots, the results do not seem to stem from publication biases (which may only have played a role with unfilled pauses and pitch). The reason for this may lie in the nature of designs of deception studies. Most deception studies investigate a number of dependent variables rather than a single dependent variable. Nonetheless, the results for all dependent variables investigated are usually reported, even if there are nonsignificant results among them. In contrast, in other research domains, studies often scrutinize effects on a single particular dependent variable. If the result of such an experiment turns out to be nonsignificant, there seems to be nothing interesting left to publish, leading such studies to vanish in file drawers (Rosenthal, 1979). Nevertheless, the fail-safe numbers obtained for the combined effect sizes were small throughout the data sets, with only one fail-safe number exceeding 15 (response latency: $k = 18$, fsn $= 19.70$). Therefore, results should be regarded with caution.

Overall effects and tests of theoretical approaches

Overall analyses weighted by sample size indicated that out of nine behaviours investigated, only two were reliably positively associated with deception: pitch ($r = .101$) and response latency ($r = .106$). There were also tendencies for deceptive messages to be shorter in duration ($r = -.057$) and to contain more speech errors ($r = .059$) which were only significant in the unweighted analyses. Table 9 summarizes our findings and compares them to those of previous meta-analyses (see below). The effect sizes of these behaviours, however, were so small that their practical relevance is questionable. Also, pitch is difficult to assess 'on-line' in face-to-face interactions without technical equipment.

It should be noted that on the basis of vote-counting it would have been concluded that there should be no substantial difference between lying and truth telling with respect to response latency. Other authors have also relied on vote-counting (e.g. Vrij, 2000), drawing partially similar but also conflicting conclusions from studies (e.g. regarding speech rate, pauses and response latency). Meta-analysts agree that more sophisticated quantitative methods of integration should be preferred over vote-counting and narrative reviews (Mullen, 1989; Rosenthal, 1994).

Table 9. Mean effect sizes of paraverbal correlates of deception in the meta-analyses by Zuckerman & Driver (1985), DePaulo et al. (2003) and this Meta-analysis

			Present meta-analysis	
Variable	Zuckerman & Driver (1985)	DePaulo et al. (2003)	Weighted mean *r*	Unweighted mean *r*
Message duration	−.09*	−.01[a,b]	−.04	−.06*
Number of words	[c]	−.01[a,b]	−.01	.01
Speech rate	−.03	.03	.01	.01
Filled pauses	.26**	.00	.04	.03
Unfilled pauses	[c]	.00	−.02	.02
Pitch	.32*	.10*	.10*	.13*
Repetitions	[c]	.10*	.08	.11
Response latency	−.01	.01	.11**	.09**
Speech errors	.11*	.00	.04	.06*

Note: In Zuckerman and Driver (1985) and DePaulo et al. (2003) effect sizes *d* were presented which were transformed into *r*s for better comparison. Positive values of *r* indicate an increase in the behaviour associated with deception. [a]Number of words or duration of message.[b] When talking time is operationalized as part of the total interaction, there is a reliable association ($r = -.17$, $p < .05$). [c]Not investigated.
$^{*}p < .05$; $^{**}p < .01$.

From an applied perspective, the small effect sizes obtained can be contrasted with beliefs about paraverbal indicators of deception. Although lay people's and professionals' assumptions converge at least in the direction of the assumed changes in paraverbal behaviours (see Tables 1 and 9), the magnitude of the expected changes is much stronger than the actual changes observed (see Breuer et al., 2005, for a direct comparison of effect sizes).

With respect to the theoretical predictions of the four approaches tested (see Table 1), none of them was unequivocally supported by our results. The arousal approach predicted correctly the increase in pitch when lying but was not supported with respect to response latency. However, moderator analyses revealed some support in studies with higher motivation (an increase in speech rate) and without sanctioning (increase in filled pauses). There was little support for the affective approach (only in the unweighted analysis of message duration: $r = -.057$). This is not surprising considering that different predictions are made depending on the type of emotion involved (Ekman, 1992) but studies did not differentiate between different types of emotions (DePaulo et al., 2003).

The attempted control approach predicted a decrease in pitch and response latency, but both behaviours increased significantly ($r = 0.101$ and $r = .106$, respectively). While the attempted control approach may have some merit in predicting a decrease in certain nonverbal behaviours associated with lying (e.g. hand and finger movements; e.g. Vrij, 1999), it largely fails to account for differences in paraverbal aspects of speech. This points to the importance of Ekman and Friesen's (1969) leakage hierarchy hypothesis that holds that some communication channels are more controllable (speech) than others (nonverbal behaviours).

The working memory model predicted several paraverbal correlates of deception correctly. Lies were told with longer response latencies, and were told with somewhat shorter message duration (significant only for the unweighted analysis). Lies also tended to contain more speech errors (again significant only for the unweighted analysis). However, the predictions of the working memory model regarding speech rate and filled and unfilled pauses were not supported by the overall analyses. The working memory model fares best when there is little opportunity for preparation and planning (see below).

As Table 9 shows, the present overall results are largely comparable to those obtained by Zuckerman et al. (1981) and Zuckerman and Driver (1985), and to those by DePaulo et al. (2003). The effect sizes for pitch and filled pauses are no longer as strong as once believed. The discrepancy regarding response latency in our and DePaulo et al.'s meta-analysis may be due to different sampling of studies as well as different estimation procedures of effect sizes when data were insufficiently reported.

Out of an array of paraverbal behaviours that were also investigated here, Zuckerman et al. (1981) as well as Zuckerman and Driver (1985) had found speech errors, speech hesitations (filled pauses), and pitch associated with deception. In addition, message duration was shorter for lies. DePaulo et al. (2003) had also observed lies to be shorter (but only with respect to talking time: $r = -.17$), higher in pitch, and containing more repetitions. Although we found an association similar in magnitude to the one by DePaulo et al. for repetitions, this association was not significant here.

Moderator Variables

From an applied perspective, the overall findings tell little without considering the impact of moderator variables investigated. Zuckerman and Driver (1985) had already pointed to the importance of motivation as well as planning in their analyses.[4] We had assumed that breaking down the research domain with respect to deception content, viz. lies concerning facts versus feelings and attitudes, would render quite different results, particularly with respect to practical implications. Unfortunately, the originally planned categorization in studies about facts only, feelings only, and facts and feelings could not be upheld due to insufficient block sizes. Hence, the comparison between studies about facts only versus those about feelings and facts combined was a compromise, which did not yield many illuminating findings (see Table 4). The association of deception with an increase in pitch and longer response latencies was only evident in studies about feelings and facts but not significant in those about facts only. As expected, messages were shorter in duration when facts were reported than when the content was both feelings and facts. Unfortunately, there were not enough studies about complex factual events in our data base, which might allow more meaningful distinctions. In future research, the distinction between lies about simple versus complex events should be paid attention to when analysing paraverbal behaviours in applied contexts. This may be particularly important for studies investigating content aspects of deception (see Sporer, 1997, 2004; Vrij, 2000).

We had also suggested (on the basis of the working memory model) that, compared to poorly prepared liars, well-prepared lies should be characterized by longer message durations, a greater number of words, higher speech rates, but fewer speech errors and shorter response latencies. This hypothesis was corroborated only with respect to response latency and message duration. Under conditions of short preparation time participants showed longer response latencies ($r = .190$) during deception, whereas with medium preparation time participants tended to respond faster when lying ($r = -.044$). Interestingly, when prepared to lie, participants did not display the typical increase in pitch during lying. Presumably, the opportunity to prepare one's lie reduced the arousal experienced by the liar.

Although the other differences were not significant, there were changes in directions of associations between behaviours observed after short versus medium preparation. As one would expect from the working memory model, message duration, filled and unfilled pauses, pitch, and response latency tended to show positive associations with deception in short preparation situations, while after medium preparation these behaviours tended to decrease when lying. However, contrary to expectations, this pattern was also observed for speech rate where a somewhat faster speech rate was observed while deceiving with little opportunity to prepare, and a somewhat slower speech rate after medium preparation.

Zuckerman and Driver (1985) had obtained contrary results with respect to speech rate. In their analysis, speech rate increased during deception under high levels of planning ($r = .29$; as was predicted from the cognitive approach), showed a nonsignificant negative association under medium ($r = -.12$), and a significant negative association under short preparation ($r = -.38$). On the other hand, we found that speech rate was slightly positively related to deception after short preparation ($r = .082$), and unrelated after medium preparation time ($r = -.034$). These difference between our and Zuckerman and Driver's analysis may either have come about by our inclusion of more recent studies, or,

[4]In their synthesis, each dependent variable was subjected to moderator analyses, without establishing whether or not the obtained effect sizes for the dependent variables were to be considered homogeneous or not. That is, Zuckerman and Driver attempted to account for variability, even when there may not have been substantial variability within the effect sizes of a dependent variable.

perhaps, by our collapsing of medium and high preparation categories which may have obscured some of the differences.

DePaulo et al. (2003) had addressed planning as a moderator only with respect to studies that had actively manipulated level of planning, resulting in a much smaller database for comparisons than our meta-analysis because we coded all studies with respect to the presumed level of preparation available to participants. These authors had found significantly longer response latencies for unplanned deceptive messages, thus confirming our conclusions, and a nonsignificant trend towards longer statements after more extensive planning. For the other behaviours examined here, the authors did not report any comparisons.

One of the most puzzling results obtained by Zuckerman and Driver (1985) was 'replicated' here: Message duration decreased during lying, when individuals were medium prepared. Zuckerman and Driver cautiously suggested that possibly 'preparing for deception intensified the anxiety of the would-be deceiver' (Zuckerman & Driver, 1985, p. 142), resulting in an increase of arousal-related behaviours. Another explanation could be that prepared deceivers concentrate so much on their before-constructed arguments that they do not need to digress, or simply have no cognitive capacity left over to digress. This may be compared with an oral examination: If the candidate knows the correct answer, he or she can do without beating about the bush.

With respect to motivation, mean effect sizes differed significantly for pitch, response latency and speech rate (see Table 6). Generally, the paraverbal behaviours were related to deception more strongly in highly motivating situations (pitch, message duration, speech rate and response latency). As one would expect from the arousal approach, highly motivated liars talked with a higher pitch in contrast to less motivated liars. Pitch seems to be difficult to control, and induced stress is likely to increase fundamental frequency (Zuckerman et al., 1981; see also Ekman, 1992; DePaulo et al., 2003). The higher speech rate may also have been a function of excitement (Ekman, 1992). However, there were no reliable differences in filled and unfilled pauses or speech errors (similar to DePaulo et al., 2003) as one would also have expected from the arousal approach. In general, any comparison between our meta-analysis and the previous ones by Zuckerman and Driver (1985) and DePaulo et al. (2003) with respect to the motivation variable is difficult due to our collapsing different levels of motivation into low and high motivation categories. Thus, the same studies may not have ended up in the same 'low' and 'high' motivation categories.

In line with the working memory model, and contrary to the attempted control approach, response latencies were reliably longer under conditions of high motivation compared to less motivated liars. The results on response latency differed from those by Zuckerman and Driver (1985) and DePaulo et al. (2003) who found no differences in the response latencies of low versus highly motivated liars.

Deceptive messages were reliably shorter in duration under high motivation but not under low motivation. The present results for message duration were similar to the results obtained by Zuckerman and Driver (1985) for response length but differed from those by DePaulo et al. (2003) who reported no differences as a function of motivation. However, in both meta-analyses response length was not separated into message duration and number of words as we had done here, so no direct comparison is possible. DePaulo et al. (2003) had found an overall decrease in talking time, however, that was not moderated by level of motivation. Perhaps, under high motivation, liars are not willing to talk as long in order not to contradict themselves. As the effect sizes for number of words had been homogeneous, we had not conducted any moderator variable analysis with this variable, which could have corroborated this hunch.

Furthermore, speech rate was significantly positively associated with deception in the high motivation condition in our study, whereas Zuckerman and Driver (1985) had found that it was negatively associated with deception ($r = -.19$). DePaulo et al. (2003) had observed no differences between low and high motivation conditions. Again, these differences between the meta-analyses are probably a function of the way studies have been categorized. But they may also reflect the different types of motivation researchers have employed (see DePaulo et al., 2003, for a thorough discussion).

We have suggested that telling lies not sanctioned by the experimenter should be associated with high motivation to lie successfully (Miller & Stiff, 1993). This association may be suitable to explain the puzzling results on speech rate. In the unsanctioned lie condition (see Table 7), speech rate increased significantly under deception. These $k = 2$ studies contributed to the shift towards the significant positive r value in the high motivation condition. In the overall analysis, no significant effect was obtained with speech rate. Apparently, under unsanctioned lie conditions, individuals seem no longer able to control the auditory channel. They talk fast and appear hasty, although according to the control approach they presumably try to speak in a calm manner. The close relationship between unsanctioned lying and highly motivated lying became obvious with respect to message duration and response latency. Effects are similarly strong in the unsanctioned lie and the high motivation conditions. It should be noted that the results on sanctioned versus unsanctioned lying must be interpreted with caution because in the data base for the unsanctioned lying condition there were often only $k = 2$ tests, and no more than $k = 4$ tests. Despite this, lying without the explicit permission of the experimenter obviously makes a difference with respect to paraverbal variables.

Other methodological aspects may play an important role as well when trying to interpret the practical implications of deception research. However, contrary to the expectation of some authors (e.g. Buller & Burgoon, 1996), the type of interaction in the experimental paradigms did not seem to make much of a difference. Surprisingly, some paraverbal associates were more likely associated with deception when they had been investigated in between-subjects designs (message duration, speech rate, filled pauses; see Table 8) while response latency only became apparent as an indicator of deception when it was studied with participants serving as their own control in within-subjects designs. Finally, it may make a difference whether or not some paraverbal behaviours are measured as duration, frequency, via ratings or with electronical devices. However, the number of studies in these comparisons are too small to draw any far-reaching conclusions. Direct experimental comparisons between different operationalizations would be necessary to draw causal conclusions which are not possible on the basis of a meta-analysis. Such comparisons would be particularly important for practical applications, especially for the attempt to create deception detection trainings for professionals using paraverbal cues.

In future research, particularly in applied settings, we think it is important to pay closer attention to the differences in operationalizations in message duration, talking time (as part of the overall interaction), message length (number of words) which may differ as a function of the complexity of the event reported and the type of interview and questions (free report, open and specific questions). In forensic settings, it is probably also important to differentiate between statements about verifiable facts—which investigators can establish independently from other sources (e.g. meteorological reports, other witnesses)—and nonverifiable contents of a message.

Liars may hold back on verifiable aspects or central details with respect to the issue investigated in order not to be caught by contradictions. On the other hand, liars may elaborate on peripheral details when constructing a lie (e.g. about an alibi) by drawing on a similar true experience. These different strategies have implications on message duration and length as well as response latencies and other aspects of speech fluency.

CONCLUSIONS

Our attempt to test the rival theoretical approaches against each other was only partially successful. However, it seems not to be sufficient to only combine these ideas. Instead one needs to specify in which situations the predictions of one approach are superior to that of another one. We have favoured a working memory model, which may be particularly useful in studying complex lies about facts (see also Sporer & Zander, 2001; Sporer et al., 2004) but may fall short when trying to explain lies about feelings and attitudes.

Nonetheless, in the context of this meta-analysis which spans many different paradigms it is difficult to establish simple hypotheses or predictions because the behavioural displays are strongly influenced by the nature of the response itself. In some situations a lie may require a simple yes/no response, whereas a truthful communication may require a narrative response. Of course, in such cases the cognitive demands involved could be higher with the truthful response (Greene et al., 1985). Additionally, the preparation time plays an important role. If someone is prepared to lie on a certain issue, but is busy with other tasks between the preparation and the deceptive response, he or she is distracted from the deceptive task. Cognitive psychologists then would expect greater difficulty in assembling the deceptive response compared to non-distracted individuals, resulting in longer response latencies (Cody et al., 1984). On the other hand, people truthfully reporting events from the distinct past may need time to reconstruct these episodic/autobiographical memories (cf. Sporer, 2004), which may slow down response latencies. Nonetheless, response latencies should be shorter when drawing on autobiographical experiences than when drawing on event scripts and having to simultaneously monitor consistency with previous statements and with objectively known facts. To sum up, the working memory approach to lie production turns out to be a very complex and multifaceted attempt at predicting paraverbal cues when deceiving but appears most promising for future research.

The present analysis of moderator variables indicated that there are considerable differences in behaviour when individuals lie with or without permission and when they are motivated to deceive successfully or not. If we start collecting different lie situations and their properties, we should be able to predict behaviour during deception more appropriately. For example, there is the well known ('white') lie towards a host, how nice and amusing a party was (although it was not). This represents a lie for reasons of politeness, which usually fades from memory of both communicator and recipient only a few minutes later. Its components are among other things that a white lie is in a way sanctioned by our society, it affords moderate cognitive effort, and its purpose is the attempt to maintain positive relationships with others. DePaulo, Kashy, Kirkendol, Wyer, and Epstein (1996) assumed that everyday lies represent the claiming of desired identities, and additionally the support of other people's claims of desired identities. This perspective is closely related to social psychological research on the self and its presentation toward others (DePaulo et al., 1996).

These 'harmless' lies should differ in their properties compared with attempts to deceive others for monetary reasons or in the courtroom. Therefore, it is likely that different types of lies also should result in different behavioural adaptations while performing these lies. For example, the present data suggest that an increase in pitch is reliably associated with deception especially when lying about feelings, without preparation and when experiencing a high motivation to lie successfully. To be of practical relevance, however, further examination of this successful cue to deception is needed because the present data suggest that pitch is positively associated with deception only when it is measured via electronical devices, whereas a negative association between deception and pitch was observed in the study by Fiedler and Walka (1993), where pitch was rated by human coders.

In other deceptive situations than the aforementioned, other cues may be more helpful. Message duration, for example, decreases reliably in deceptive situations involving factual statements ($r = -.075$), opportunity to prepare the lie ($r = -.068$), and a high motivation to lie successfully ($r = -.102$). What we probably need is a taxonomy of situations for which different theoretical approaches are applicable. It is a challenge to create experimental situations in which different conditions of lying are present. In particular, relatively little is still known about high-stake lies. As depicted in Table 7, only very few researchers were successful in or even thought about creating unsanctioned lie conditions (Buller, Burgoon, Buslig, & Roiger, 1994; deTurck & Miller, 1985; Exline, Thibaut, Hickey, & Gumpert, 1971; Feeley & deTurck, 1998; Mehrabian, 1971; Stiff & Miller, 1986).

In recent research, however, there were attempts to move away from investigating deceptive behaviour via laboratory studies only. Vrij and Mann (2001) investigated the behaviour of a real suspect while he was interviewed by the police in a murder case, separating interview parts for which there was corroborating and disconfirming evidence. Similarly, Davis and Hadiks (1995) analysed an interview with Saddam Hussein by an American journalist. That is, these authors chose nonexperimental paradigms with post hoc assessment of the truthful versus deceptive utterances (of the murderer, or of Saddam Hussein, respectively). Davis and Hadiks (1995) argued that in past research, there had been too much emphasis on 'general categories and simplistic coding tasks' (Davis & Hadiks, 1995, p. 6). They also assumed that cues to deception must be context-bound, as described above. Possibly, the too general and simple rules of rating behaviours established in deception experiments are the reason why the majority of behaviours failed to distinguish between liars and truth tellers. Paraverbal behaviours such as message length and number of words may clearly depend upon whether or not specific aspects of a message can be independently verified by an interaction partner. In case this is possible, communicators may withhold information not to be caught by evidence to the contrary but also to minimize working memory load to avoid inconsistencies with earlier utterances. However, individual differences in deceptive skills need also be considered as people may favor different forms of 'strategic communication'.

Other researchers have attempted to analyse not only nonverbal and paraverbal correlates of deception, but also the content of messages themselves (e.g. Koehnken, 1990; Mehrabian, 1971; Sporer, 1997, 2004; Vrij, 2000). In fact, the recent comprehensive meta-analysis by DePaulo et al. (2003) indicates that some of the content cues to deception—which are better described as 'truth criteria'—show larger effect sizes than most of the nonverbal and paraverbal behaviours investigated (for reviews, see Masip, Sporer, Garrido, & Herrero, 2005; Sporer, 2004; Vrij, 2000).

Although the thorough analysis of an individual's behaviour seems to be a challenging step within the arena of deception research, these methods often do not seem applicable in face-to-face interactions, when an individual wishes to assess credibility of his or her counterpart immediately. Analyses of pitch for example are only possible using sophisticated technical equipment and investing very much time. Therefore, they are not suitable to meet the goal of identifying reliable deception cues in order to train professionals whose work depends on assessing credibility 'on the spot'.

ACKNOWLEDGEMENTS

This research was supported by a grant from the Deutsche Forschungsgemeinschaft (Sp 262/3-2) to the first author. We are indebted to Jennifer Zander for her help in coding some of the data.

REFERENCES*

References marked with an asterisk (*) were used for the meta-analysis.

Akehurst, L., Koehnken, G., Vrij, A., & Bull, R. (1996). Lay persons' and police officers' beliefs regarding deceptive behavior. *Applied Cognitive Psychology, 10*, 461–471.

Anderson, S. J., & Conway, M. A. (1997). Representation of autobiographical memories. In M. A. Conway (Ed.), *Cognitive models of memory* (pp. 217–246). Hove, Sussex, England: Psychology Press.

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: a proposed system and its control processes. In K. W. Spence (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 2, pp. 89–195). New York: Academic Press.

*Akehurst, L., & Vrij, A. (1999). Creating suspects in police interviews. *Journal of Applied Social Psychology, 29*, 192–210.

*Anolli, L., & Ciceri, R. (1997). The voice of deception: vocal strategies of naive and able liars. *Journal of Nonverbal Behavior, 21*, 259–284.

Baddeley, A. (2000a). The episodic buffer: a new component of working memory? *Trends in Cognitive Sciences, 4*, 417–423.

Baddeley, A. (2000b). Short-term and working memory. In E. Tulving, & F. I. M. Craik (Eds.), *Handbook of memory* (pp. 77–92). Oxford, UK: Oxford University Press.

Breuer, M. M., Sporer, S. L., & Reinhard, M.-A. (2005). Subjektive Indikatoren von Täuschung: die Bedeutung von Situation und Gelegenheit zur Vorbereitung [Subjective indicators of deception: the role of situational determinants and preparation]. *Zeitschrift für Sozialpsychologie, 36*, 189–201.

Brewer, W. F. (1986). What is autobiographical memory? In D. C. Rubin (Ed.), *Autobiographical memory* (pp. 25–49). Cambridge: Cambridge University Press.

Brewer, W. F. (1996). What is recollective memory? In D. C. Rubin (Ed.), *Remembering our past: Studies in autobiographical memory* (pp. 19–66). New York: Cambridge University Press.

*Bond, C. F., Nelson Kahler, K., & Paolicelli, L. M. (1985). The miscommunication of deception: an adaptive perspective. *Journal of Experimental Social Psychology, 21*, 331–345.

*Bond, C. F., Omar, A., Mahmoud, A., & Bonser, R. N. (1990). Lie detection across cultures. *Journal of Nonverbal Behavior, 14*, 189–204.

Buller, D. B., & Burgoon, J. K. (1996). Interpersonal deception theory. *Communication Theory, 6*, 203–242.

Buller, D. B., Burgoon, J. K., Buslig, A. L., & Roiger, J. F. (1994). Interpersonal deception VIII. Further analysis of nonverbal and verbal correlates of equivocation from the Bavelas et al. (1990) research. *Journal of Language and Social Psychology, 13*, 396–417.

Buller, D. B., Burgoon, J. K., White, C. H., & Ebesu, A. S. (1994). Interpersonal deception VII. behavioral profiles of falsification, equivocation, and concealment. *Journal of Language and Social Psychology, 13*, 366–395.
*Burns, J. A., & Kintz, B. L. (1976). Eye contact while lying during an interview. *Bulletin of the Psychonomic Society, 7*, 87–89.
*Cody, M. J., Lee, W., & Chao, E. Y. (1989). Telling lies: correlates of deception among Chinese. In J. P. Forgas, & J. M. Innes (Eds.), *Recent advances in Social Psychology: An international perspective* (pp. 359–368). North-Holland, Netherlands: Elsevier Science Publishers.
Cody, M. J., Marston, P. J., & Foster, M. (1984). Deception: paralinguistic and verbal leakage. In R. M. Bostrom (Ed.), *Communication yearbook 8* (pp. 464–490). Beverly Hills, CA: Sage.
Conway, M. A.(1990). *Autobiographical memory: An introduction.* Milton Keynes, UK: Open University Press.
Davis, M., & Hadiks, D. (1995). Demeanor and credibility. *Semiotica, 106*, 5–54.
DePaulo, B. M., Charlton, K., Cooper, H., Lindsay, J. J., & Muhlenbruck, L. (1997). The accuracy–confidence correlation in the detection of deception. *Personality and Social Psychology Review, 1*, 346–357.
DePaulo, B. M., Kashy, D. A., Kirkendol, S. E., Wyer, M. M., & Epstein, J. A. (1996). Lying in everyday life. *Journal of Personality and Social Psychology, 70*, 979–995.
DePaulo, B. M., Lindsay, J. J., Malone, B. E., Muhlenbruck, L., Charlton, K., & Cooper, H. (2003). Cues to deception. *Psychological Bulletin, 129*, 74–112.
*DePaulo, B. M., Rosenthal, R., Rosenkrantz, & Green, C. R. (1982). Actual and perceived cues to deception: a closer look at speech. *Basic and Applied Social Psychology, 3*, 291–312.
*DePaulo, P. J., & DePaulo, B. M. (1989). Can deception by salespersons and customers be detected through nonverbal behavioral cues? *Journal of Applied Social Psychology, 19*, 1552–1577.
*deTurck, M. A., & Miller, G. R. (1985). Deception and arousal: isolating the behavioral correlates of deception. *Human Communication Research, 12*, 181–201.
*Donaghy, W. C., & Dooley, B. (1994). Head movement, gender, and deceptive communication. *Communication Reports, 7*, 67–75.
*Ebesu, A. S., & Miller, M. D. (1994). Verbal and nonverbal behaviors as a function of deception type. *Journal of Language and Social Psychology, 13*, 418–442.
Ekman, P. (1981). Mistakes when deceiving. *Annals of the New York Academy of Sciences, 364*, 269–278.
Ekman, P. (1988). Lying and nonverbal behavior: theoretical issues and new findings. *Journal of Nonverbal Behavior, 12*, 163–176.
Ekman, P. (1992). *Telling lies: Clues to deceit in the marketplace, politics and marriage.* New York: W. W. Norton.
Ekman, P., & Friesen, W. V. (1969). Nonverbal leakage and clues to deception. *Psychiatry, 32*, 88–106.
Ekman, P., Friesen, W. V., & O'Sullivan, M. (1988). Smiles when lying. *Journal of Personality and Social Psychology, 54*, 414–420.
*Ekman, P., Friesen, W. V., & Scherer (1976). Body movement and voice pitch in deceptive interaction. *Semiotica, 16*, 23–27.
Ekman, P., & O'Sullivan, M. (1991). Who can catch a liar? *American Psychologist, 46*, 913–920.
*Ekman, P., O'Sullivan, M., Friesen, W. V., & Scherer, K. R. (1991). Face, voice, and body in detecting deceit. *Journal of Nonverbal Behavior, 15*, 125–135.
*Exline, R. E., Thibaut, J., Hickey, C. B., & Gumpert, P. (1970). Visual interaction in relation to Machiavellianism and an unethical act. In R. Christie, & F. L. Geis (Eds.), *Studies in Machiavellianism* (pp. 53–75). New York: Academic Press.
Feeley, T. H., & deTurck, M. A. (1998). The behavioral correlates of sanctioned and unsanctioned deceptive communication. *Journal of Nonverbal Behavior, 22*, 189–204.
Fiedler, K., & Walka, I. (1993). Training lie detectors to use nonverbal cues instead of global heuristics. *Human Communication Research, 20*, 199–223.
Fiske, S. T., & Taylor, S.E. (1991). *Social cognition.* London: McGraw-Hill.
Goldman-Eisler, F. (1968). *Psycholinguistics: Experiments in spontaneous speech.* New York: Doubleday.
Greene, J. O., O'Hair, D., Cody, M. J., & Yen, C. (1985). Planning and control of behavior during deception. *Human Communication Research, 11*, 335–364.

Hale, J. L., & Stiff, J. B. (1990). Nonverbal primacy in veracity judgments. *Communication Reports, 3*, 75–83.

*Harrison, A. A., Hwalek, M., Raney, D. F., & Fritz, J. G. (1978). Cues to deception in an interview situation. *Social Psychology, 41*, 156–161.

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

*Hocking, J. E., & Leathers, D. G. (1980). Nonverbal indicators of deception: a new theoretical perspective. *Communication Monographs, 47*, 119–131.

*Hoefer, E., Koehnken, G., Hanewinkel, R., & Bruhn, C. (1993). *Diagnostik und Attribution von Glaubwuerdigkeit* [Assessment and Attribution of Credibility]. Kiel: Final report to the Deutsche Forschungsgemeinschaft, KO 882/4-2. Kiel, Germany: University of Kiel.

Kahneman, D. (1973). *Attention and effort.* Englewood Cliffs, NJ: Prentice-Hall.

*Knapp, M. L., Hart, R. P., & Dennis, H. S. (1974). An exploration of deception as a communication construct. *Human Communication Research, 1*, 15–29.

Koehnken, G. (1990). *Glaubwuerdigkeit* [Credibility]. Muenchen, Germany: Psychologie Verlags Union.

*Kraut, R. E. (1978). Verbal and nonverbal cues in the perception of lying. *Journal of Personality and Social Psychology, 36*, 380–391.

*Kraut, R. E., & Poe, D. (1980). On the line: the deception judgments of customs inspectors and laymen. *Journal of Personality and Social Psychology, 39*, 784–798.

Light, R. J., & Pillemer, D. B. (1984). *Summing up: The science of reviewing research.* Cambridge, MA: Harvard University Press.

Lipsey, M. W. (1994). Identifying potentially interesting variables and analysis opportunities. In H. Cooper, & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 111–123). New York: Russel Sage Foundation.

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis.* Thousand Oaks, CA: Sage.

*McClintock, C. C., & Hunt, R. G. (1975). Nonverbal indicators of affect and deception in an interview setting. *Journal of Applied Social Psychology, 5*, 54–67.

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, and Law, 11*, 99–122.

Mehrabian, A. (1971). Nonverbal betrayal of feeling. *Journal of Experimental Research in Personality, 5*, 64–73.

*Miller, G. R. (1983). Telling it like it isn't and not telling it like it is: some thoughts on deceptive communication. In J. I. Sisco (Ed.), *The Jensen lectures: Contemporary communication studies* (pp. 91–116). Tampa: University of South Florida.

*Miller, G. R., deTurck, M. A., & Kalbfleisch, P. J. (1983). Self-monitoring, rehearsal, and deceptive communication. *Human Communication Research, 10*, 97–117.

Miller, G. R., & Stiff, J. B. (1993). *Deceptive communication.* Newbury Park, CA: Sage.

Miyake, A., & Shah, P. (Eds.). (1999). *Models of working memory: Mechanisms of active maintenance and executive control.* New York: Cambridge University Press.

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

*O'Hair, H. D., Cody, M. J., & McLaughlin, M. L. (1981). Prepared lies, spontaneous lies, Machiavellianism, and nonverbal communication. *Human Communication Research, 7*, 325–339.

*Rockwell, P., Buller, D. B., & Burgoon, J. K. (1997). Measurement of deceptive voices: comparing acoustic and perceptual data. *Applied Psycholinguistics, 18*, 471–484.

Rosenthal, R. (1979). The ''file drawer problem'' and tolerance for null results. *Psychological Bulletin, 86*, 638–641.

Rosenthal, R. (1991). *Meta-analytic procedures for social research*. Newbury Park, CA: Sage.

Rosenthal, R. (1994). Parametric measures of effect size. In H. Cooper, & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 231–244). New York: Russell Sage Foundation.

Schank, R. C., & Abelson, R. P. (1977). *Scripts, plans, goals and understanding.* Hillsdale, N.J.: Erlbaum.

*Schneider, S. M., & Kintz, B. L. (1977). The effect of lying upon foot and leg movement. *Bulletin of the Psychonomic Society, 10*, 451–453.

Sporer, S. L. (1997). The less traveled road to truth: verbal cues in deception detection in accounts of fabricated and self-experienced events. *Applied Cognitive Psychology, 11*, 373–397.

Sporer, S. L. (2004). Reality monitoring and the detection of deception. In P.-A. Granhag, & L. Stromwall (Eds.), *Deception detection in forensic contexts* (pp. 64–102). Cambridge University Press: Cambridge, UK.

Sporer, S. L., Krompass, M., & Breuer, M. M. (2005). *Paraverbal correlates of deception: Does preparation make a difference?* Manuscript submitted for publication.

Sporer, S. L., & Zander, J. (2001, June). *Nonverbal cues to deception: Do motivation and preparation make a difference?* Paper presented at the 11th European Conference on Psychology and Law in Lisbon, Portugal.

Stiff, J. B., & Miller, G. R. (1986). "Come to think of it . . . ": Interrogative probes, deceptive communication, and deception detection. *Human Communication Research, 12*, 339–357.

Stock, W. A. (1994). Systematic coding for research synthesis. In H. Cooper, & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 125–139). New York: Russell Sage Foundation.

*Streeter, L. A., Krauss, R. M., Geller, V., Olson, C., & Apple, W. (1977). Pitch changes during attempted deception. *Journal of Personality and Social Psychology, 35*, 345–350.

Stromwall, L. A., Granhag, P. A., & Hartwig, M. (2004). Practitioners' beliefs about deception. In P. A. Granhag, & L. A. Stromwall (Eds.), The detection of deception in forensic contexts (pp. 229–250). New York: Cambridge University Press.

*Todd-Mancillas, W. R., & Kibler, R. J. (1979). A test of concurrent validity for linguistic indices of deception. *Western Journal of Speech Communication, 43*, 108–122.

Tulving, E. (1989). Memory: performance, knowledge, and experience. *European Journal of Cognitive Psychology, 1*, 3–26.

Vrij, A. (1999). Behavioral correlates of deception in a simulated police interview. *Journal of Psychology, 129*, 15–28.

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

*Vrij, A., Akehurst, L., & Morris, P. (1997). Individual differences in hand movements during deception. *Journal of Nonverbal Behavior, 21*, 87–102.

Vrij, A., Dragt, A. W., & Koppelaar, L. (1992). Interviews with ethnic interviewees: nonverbal communication errors in impression formation. *Journal of Community and Applied Social Psychology, 2*, 199–209.

Vrij, A., Edward, K., & Bull, R. (2001). Police officers' ability to detect deceit: the benefit of indirect deception detection measures. *Legal and Criminological Psychology, 6*, 185–197.

*Vrij, A., Edward, K., Roberts, K. P., & Bull, R. (2000). Detecting deceit via analysis of verbal and nonverbal behavior. *Journal of Nonverbal Behavior, 24*, 239–263.

*Vrij, A., & Heaven, S. (1999). Vocal and verbal indicators of deception as a function of lie complexity. *Psychology, Crime and Law, 5*, 203–215.

Vrij, A., & Mann, S. (2001). Telling and detecting lies in a high-stake situation: the case of a convicted murderer. *Applied Cognitive Psychology, 15*, 187–203.

*Vrij, A., & Winkel, F. W. (1991). Cultural patterns in Dutch and Surinam nonverbal behavior: an analysis of simulated police/citizen encounters. *Journal of Nonverbal Behavior, 15*, 169–184.

*Vrij, A., & Winkel, F. W. (1993). Objective and subjective indicators of deception. *Issues in Criminological and Legal Psychology, 20*, 51–57.

Wheeler, M. A., Stuss, D. T., & Tulving, E. (1997). Toward a theory of episodic memory: the frontal lobes and autonoetic consciousness. *Psychological Bulletin, 121*, 331–354.

Winer, B. J., Brown, D. R., & Michels, K. M. (1991). *Statistical principles in experimental design* (3rd edn). New York: McGraw-Hill.

Zuckerman, M., Amidon, M. D., Bishop, S. E., & Pomerantz, S. D. (1982). Face and tone of voice in communication of deception. *Journal of Personality and Social Psychology, 43*, 347–357.

*Zuckerman, M., deFrank, R. S., Hall, J., Larrance, D. T., & Rosenthal, R. (1979). Facial and vocal cues of deception and honesty. *Journal of Experimental Social Psychology, 15*, 378–396.

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. 1–59). New York: Academic Press.

Zuckerman, M., & Driver, R. (1985). Telling lies: verbal and nonverbal correlates of deception. In A. W. Siegman, & S. Feldstein (Eds.), *Nonverbal communication: An integrated perspective* (pp. 129–147). Hillsdale, NJ: Lawrence Erlbaum.

