



How People *Really* Suspect Lies: A Re-examination of Novotny et al.'s (2018) Data

Jaume Masip¹ · Nuria Sánchez¹

Published online: 21 June 2019
© Springer Science+Business Media, LLC, part of Springer Nature 2019

Abstract

Previous research has shown that in real-life situations people detect lies mostly from non-behavioral information (e.g., physical evidence, third-party information, confessions...) rather than from behavioral cues. Novotny et al. (J Nonverbal Behav 42:41–52, 2018, <https://doi.org/10.1007/s10919-017-0263-2>) argued that while lies are *detected* primarily from non-behavioral information, initial *suspicion* of a lie can be triggered primarily from behavioral cues. They conducted two studies and claimed support for their hypotheses. However, there are a number of problematic issues with Novotny et al.'s research and conclusions. We conducted analyses based on the frequencies and percentages they reported, and used meta-analytical techniques to combine their findings concerning discovered lies with those of previous research. The results show that lies are indeed detected from non-behavioral information more often than from behavioral cues. However, contrary to Novotny et al.'s assertions, suspicion is *not* triggered primarily from behavioral cues—rather, there is a trend in favor of non-behavioral information. Even so, behavioral cues play a bigger role in eliciting suspicion than in lie discovery.

Keywords Deception · Lie detection · Deception cues · Suspicion · Behavioral cues

Introduction

Deception detection research has been mostly conducted in laboratory settings. In a typical laboratory experiment, participants (normally college students) are randomly assigned to either lie or tell the truth. The participants' truthful or deceptive messages are typically video recorded and shown to unacquainted observers who assess the senders' veracity. The words and nonverbal behavior of the senders while lying and telling the truth can also be coded for analysis. Two main conclusions of this area of research are that most behavioral cues to deception are faint and unreliable (DePaulo et al. 2003) and that, consequently, people are rather inaccurate in judging veracity from behavioral cues [54% average accuracy, just above 50% chance accuracy; see Bond and DePaulo (2006)].

✉ Jaume Masip
jmasip@usal.es

¹ Department of Social Psychology and Anthropology, Psychology School, University of Salamanca, Avenida de la Merced, 109-131, 37005 Salamanca, Spain

Table 1 Basic design of Novotny et al.'s (2018) Studies

Condition	Dependent measures	
	Behavioral cues	Non-behavioral information
Discovered lie	a	b
Suspected lie	c	d

However, it has been argued that the typical laboratory experiment fails to mirror the circumstances surrounding real-life deception and its detection (Levine 2018). For example, unlike what happens in laboratory experiments, in real life people might become wary of the honesty of acquainted others (as opposed to strangers) and can therefore have access to highly diagnostic non-behavioral information (as opposed to unreliable behavioral cues). These notions were anticipated by Park et al. (2002), who asked college students to recall a lie they had detected in the past and to report how they had detected it. They found that lies were primarily discovered from non-behavioral information such as physical evidence, solicited or unsolicited confessions, third-party information, or inconsistencies with knowledge.¹ More recently, Masip and Herrero (2015) replicated Park et al.'s main results with both community members and police officers.

In a recent article, Novotny et al. (2018) highlighted the distinction between *suspecting* a lie and *discovering* a lie, and contended that although only strong, undisputed non-behavioral information can “prove” a lie (discovery), behavioral cues can nevertheless act as red flags triggering suspicion. Novotny et al. suggested lie detection is a two-step process wherein behavioral cues elicit suspicion, and then confirmation is sought via more compelling non-behavioral information.

To test their ideas, Novotny et al. (2018) conducted two studies. Their paradigm was similar to Park et al.'s (2002), but Novotny et al. asked some of their participants to recall a lie they had *suspected*, and some others to recall a lie they had *discovered*. The basic design of Novotny et al.'s studies is displayed in Table 1, where each cell is named with a letter. Table 2 contains Novotny et al.'s hypotheses, as well as the cells (denoted by Table 1's letters) being compared.

Unfortunately, Novotny et al.'s (2018) article suffers from a number of problems. One such problem is that in some cases the authors claimed support for specific hypotheses by just looking at frequencies and/or percentages without reporting the outcome of any formal statistical test. However, our major concern (which motivated the current paper) refers to the way Novotny et al. tested H1b (see Table 2), as well as to their inaccurate conclusions regarding this hypothesis. The outcomes of Novotny et al.'s studies are displayed in Table 3. In view of these outcomes, statements such as “participants who were asked about suspicion were more likely to report nonverbal/verbal behavior [cell c] than harder evidence [cell d]” (Novotny et al. 2018, p. 47) are puzzling and undoubtedly inaccurate. The

¹ Behavioral cues are visible, vocal, and verbal potential cues to deception displayed by the liar at the time the lie is told [i.e., the kinds of cues typically examined by laboratory research on deception detection; see DePaulo et al. (2003)]. Non-behavioral cues involve information about the context (i.e., information that goes beyond the liar's behavior during the specific deceptive exchange) that can contradict the liar's statement, such as third-party information, specific knowledge of the detector, physical evidence, or the liars' ultimate admission that the exchange was deceptive.

Table 2 Hypotheses of Novotny et al.'s (2018) studies

Study	Hypotheses	Cell comparison
Study 1	H1a. "Participants will report non-behavioral evidence ... more often than behavioral evidence in the discovered condition" (p. 42)	b > a
	H1b. "Participants will report behavioral evidence more often than non-behavioral evidence in the suspected condition" (p. 42)	c > d
	H2a. "Participants will report significantly more non-behavioral evidence than behavioral evidence when asked about discovering versus suspecting a lie." (p. 42)	b > d
	H2b. "Participants will report significantly more behavioral evidence than non-behavioral evidence when asked about suspecting versus discovering a lie." (p. 42)	c > a
Study 2	H1a. "Participants will report non-behavioral evidence as more important than behavioral evidence in the discovered condition." (p. 47)	b > a
	H1b. "Participants will report behavioral evidence as more important than non-behavioral evidence in the suspected condition." (p. 47)	c > d

The phrasing of H2a and H2b looks confusing to us, but in reporting the results the authors emphasize the comparisons indicated in the final column

Table 3 Outcomes of Novotny et al.'s (2018) studies

Study	Condition	Dependent measures	
		Behavioral cues (%)	Non-behavioral information (%)
Study 1	Discovered lie	1.3–5.3	98.7
	Suspected lie	43.4–54.2	56.6
Study 2	Discovered lie	14.0	86.0
	Suspected lie	37.6	62.4

For Study 1 the numbers are percentages of *lies* detected/suspected from behavioral cues or non-behavioral information. For Study 2 the numbers are percentages of *cues* (either behavioral or non-behavioral) leading participants to detect/suspect deception

goals of this paper are first to highlight and then suggest fixes to Novotny et al.'s errors and misguided conclusions. To do so, we needed to know the specific percentages of behavioral and non-behavioral cues in each condition. This also proved to be problematic, as the way Novotny et al. calculated the number of cues in Study 1 makes it impossible to know for certain how many behavioral cues were mentioned by the participants. However, we were able to establish a range, which we used for our calculations.

In the following, we first describe the difficulties we faced to calculate the exact frequencies/percentages of behavioral cues in Study 1 and how we arrived at the estimated range. Next we turn to hypothesis testing, including the testing of H1b. Just after that, we focus on the substantive issue of Novotny et al.'s counter-factual statements relative to H1b. Our ultimate goal is to clarify what the conclusions of Novotny et al.'s (2018) research are. To have a more accurate picture of the strength of the evidence, we also used meta-analytical techniques to combine Novotny et al.'s data with those of Park et al. (2002) and Masip and Herrero (2015).

Questionable Aspects in Novotny et al.'s (2018) Article

Frequency/Percentage Calculations

For the purposes of the current paper, we needed to ascertain the precise frequencies and percentages of behavioral cues and non-behavioral information in Novotny et al.'s (2018) studies. Novotny et al. differentiated between seven cue categories that were grouped as (a) "at the time verbal/nonverbal cues" (i.e., behavioral cues, one category), (b) non-behavioral information (five cue categories such as third-party information, physical evidence, etc.), and (c) a "combination of factors." This latter category was used when the participant mentioned several cues that pertained to more than one of the other six cue categories. Those cues could either be all non-behavioral or a combination of non-behavioral and behavioral cues. In Study 1, the authors reported the percentage of *lies* that had been suspected or discovered on the basis of each of the seven categories—i.e., including the combined category. This involves uncertainty concerning the precise frequency (or percentage) of behavioral cues, as we don't know how many of the combined-category cases involved a behavioral cue. It is worth mentioning, however, that all combined-category cases involved at least one non-behavioral cue, because the combined category included cues of more than

one category and there was only one category of behavioral cues (compared to five categories of non-behavioral cues).

In Study 2, the authors addressed this issue by counting how many cues pertaining to each of the previous six categories were included in the “combination of factors” category. Then they reported the percentage of cues belonging to each of the six non-behavioral and behavioral cue categories. Having used two different procedures to calculate percentages (Study 1: percentage of lies, and combined category as a separate category; Study 2: percentage of cues, and combined category broken down into separate behavioral and non-behavioral categories) makes it difficult to compare the results of the two studies. Further, as noted earlier, the procedure used in Study 1 creates uncertainty concerning the measures of interest.

Specifically, in the *discovery* condition of Study 1, 1.3% of lies were discovered from behavioral cues alone, 94.7% from non-behavioral information alone, and 4.0% from a combination of cues. Because the combination category necessarily involved non-behavioral information, the total percentage for non-behavioral information was $94.7\% + 4.0\% = 98.7\%$ (not 96%, which is the figure provided by Novotny et al.; it is unclear to us how they arrived at this latter percentage). Because we don't know how many of the 4.0% cases in the combination of cues category involved behavioral cues, the actual percentage for behavioral cues fell somewhere between 1.3 and 5.3% ($1.3\% + 4.0\%$). Concerning the *suspected-lie* condition, behavioral cues alone elicited suspicion in 43.4% of cases, non-behavioral information alone in 45.8% of cases, and a combination of cues in 10.8% of cases. Again, because the authors offered no clue allowing readers to know how many of those 10.8% of cases involved behavioral cues, the actual percentage of lies suspected on the basis of non-behavioral cues was 56.6% ($45.8\% + 10.8\%$), while the actual percentage of lies suspected on the basis of behavioral cues could range between 43.4 and 54.2% ($43.4\% + 10.8\%$). Those percentages are displayed in the upper panel of Table 3. The results of Study 2 are in the lower panel of Table 3. However, note that while for Study 1 the numbers are the percentage of lies detected/suspected from each kind of information (behavioral/non-behavioral), for Study 2 they reflect the percentages of behavioral/non-behavioral cues leading participants to detect/suspect deception.

Hypothesis Testing

We used the numbers in Table 3 to formally test Novotny et al.'s (2018) hypotheses (see Tables 1, 2). We conducted these analyses because in some cases Novotny et al. made their claims without reporting the outcome of any formal inference test, and in some other cases the tests they conducted may not be the most appropriate for the data at hand. We first focus on H1a (b > a comparison), H2a (b > d comparison), and H2b (c > a comparison; see Tables 1, 2). Our analyses provided support for all these hypotheses. We tested H1a in Study 1 with a McNemar test² using the most conservative (in terms of finding support for H1a) percentage for behavioral cues (i.e., 5.3%; see Table 3). The difference between the percentages for behavioral cues and non-behavioral information was -93.33 , 95% CI $[-97.23, -81.71]$, with $\chi^2_{\text{uncorrected}} = 68.06$, $p < .001$, and $\chi^2_{\text{corrected}} = 66.13$, $p < .001$. It is

² The McNemar tests were run in SPSS using Marta García-Granero's macro, which is available, along with instructions on how to use it, at <http://www.how2stats.net/2011/09/two-proportions-test-related-spss.html>. The confidence intervals for the difference between proportions are based on a procedure described by Newcombe (1998). The Chi square correction used was the conservative Yates' correction.

Table 4 Outcomes of the Z tests for two population proportions conducted on Novotny et al.'s (2018) data

Study	Hypothesis	Z	<i>p</i> (two tailed)
Study 1	H2a	6.23	< .001
	H2b ^a	−5.49	< .001
Study 2	H2a	3.63	< .001
	H2b	−3.63	< .001

^aTested with the most conservative percentages in terms of finding support for the hypothesis (i.e., 5.3% for the discovered-lie condition and 43.4% for the suspected lie condition)

Table 5 Standardized residuals, Chi square values, and associated *p* values for the associations between condition and whether behavioral and non-behavioral cues were (“Yes”) or were not (“No”) mentioned by the participants in Novotny et al.'s (2018) Studies 1 and 2

Study	Condition	Behavioral		Non-Behavioral	
		Yes	No	Yes	No
Study 1	Discovered lie	−3.99	2.68	2.19	−3.95
	Suspected lie	3.80	−2.55	−2.08	3.76
	χ^2 (1)	44.01		38.83	
	<i>p</i>	< .001		< .001	
Study 2	Discovered lie	−2.17	1.26	1.26	−2.17
	Suspected lie	2.27	−1.32	−1.32	2.27
	χ^2 (1)	13.17		13.17	
	<i>p</i>	< .001		< .001	

clear, therefore, that in the discovered-lie condition non-behavioral information was indeed mentioned significantly more often than behavioral cues. Unfortunately, not enough information was provided by Novotny et al. to run a similar test for Study 2, although the percentages (see Table 3) suggest that it is very likely that H1a was also supported for Study 2.

Novotny et al.'s wording of H2a and H2b was quite confusing (see Table 2), but the results sections of their paper helped clarify these hypotheses. We used Z tests for two population proportions to test them. The outcomes are shown in Table 4. We also conducted two 2×2 Chi square tests for each study to examine the associations between condition (discovered vs. suspected lie) and whether behavioral cues (first test) and non-behavioral information (second test) were mentioned or not. All four Chi square values were significant, and critical standardized residuals were larger than 1.96 (the residuals for Study 2—mentioning non-behavioral information being the only exception; Table 5). Both the Z tests and the Chi square tests suggest that in both studies significantly more non-behavioral information (H2a) and fewer behavioral cues (H2b) were mentioned in the discovered-lie than in the suspected-lie condition.

Our major concerns (and our main motivation to write this paper) refer to the way Novotny et al. (2018) tested H1b and their faulty conclusions concerning this hypothesis. H1b predicted that $c > d$ (see Tables 1, 2). However, we conducted McNemar tests that failed to reveal any significant difference between the percentage of behavioral and non-behavioral cues in the suspected-lie condition in Study 1; considering the 43.4% rate of behavioral cues: difference between proportions = −13.25, 95% CI [−33.52, 8.19], $\chi^2_{\text{uncorrected}} = 1.46$, $p = .227$, $\chi^2_{\text{corrected}} = 1.21$, $p = .272$; considering the 54.2% rate of behavioral cues: difference between proportions = −2.41, 95% CI [−22.17, 17.55], $\chi^2_{\text{uncorrected}} = 0.05$, $p = .816$, $\chi^2_{\text{corrected}} = 0.01$, $p = .908$. In fact, the direction of the effect was contrary to H1b; that is,

the percentage was higher for non-behavioral information than for behavioral cues. The authors do not report enough data to conduct the same test for Study 2, but in Study 2 the difference in proportions was even larger than it was in Study 1 (and likely significant) and, again, contrary to H1b (see Table 3). Thus, definitely, H1b was *not* supported in either study.

Misguided Conclusions in Novotny et al.'s (2018) Paper

This latter evidence is at odds with Novotny et al.'s conclusions regarding H1b. As noted above, they stated (relative to Study 1) that “participants who were asked about suspicion were more likely to report nonverbal/verbal behavior than harder evidence” (p. 47). In view of the data analyses described above, it is clear that this statement is false. Furthermore, in presenting the results of Study 1, Novotny et al. wrote that “*the single most important factor* when participants were asked about suspected lies was the behavioral evidence, *as predicted in H1b*” (p. 46, italics added). However, it is clear (see Table 2) that H1b stated nothing concerning “the single most important factor.” Therefore, Novotny et al.'s statement is indeed misleading.

Concerning Study 2, on p. 49 the authors wrote: “consistent with H1b, participants in suspected conditions reported using behavioral evidence significantly more often than those in discovered conditions” (p. 49). The same statement (including the reference to H1b) is made again at the end of the Study discussion. However, note that those statements refer to the *a* versus *c* comparison (see Table 1), not the *c* versus *d* comparison. Therefore, they are fully unrelated to H1b. However, they can mislead the reader into believing that H1b was factually supported by the data. It was not.

We consider those statements serious distortions of the actual findings. Practices of this kind have an enormous potential to mislead readers. Their impact on scientific knowledge is extremely negative. Our purpose in writing this article was to remedy this problem of Novotny et al.'s paper. However, rather than to just highlight the problematic issues, our goal is to also emphasize the positive contribution of Novotny et al.'s actual findings to the lie detection literature. The next section focuses on the real significance of Novotny et al.'s contribution.

What Do the Novotny et al.'s Data Actually Tell Us?

We believe that despite H1b not being supported by the data, the main argument put forward by Novotny et al. (2018) still holds. Their main point is that while the impact of behavioral cues on the discovery of deception is limited, behavioral cues still play a role in eliciting suspicion. But note that for this prediction to be supported, it is not necessary that significantly more behavioral than non-behavioral information is mentioned by participants in the suspected-lie condition (H1b). It suffices that the likelihood of mentioning behavioral (relative to non-behavioral) information is higher in the suspected-lie condition than in the discovered-lie condition. And this is exactly what the data show.

This is already apparent from the observation of Table 5, but can also be demonstrated with effect sizes such as Odds Ratio (*OR*) or Risk Difference (*RD*), which are typically used to estimate risk in medical research (see Ferguson 2009; Lipsey and Wilson 2001). Using the cell notation in Table 1, $OR = (c/d)/(a/b)$. Note that if the odds of mentioning behavioral cues are the same both in the discovered-lie and the suspected-lie conditions, then *OR*

equals 1. However, in Novotny et al.'s Study 1, $OR=17.7$, and in Study 2 $OR=3.7$.³ This indicates that the odds of mentioning behavioral relative to non-behavioral information are much greater in the suspected-lie condition compared to the discovered-lie condition (as benchmarks, let us consider that according to Ferguson (2009) $OR=2.0$ is a small effect, $OR=3.0$ is a moderate effect, and $OR=4.0$ is a strong effect).

RD is easier to interpret than OR (Ferguson 2009). To calculate RD in the current case, we need to first consider what proportion of all cues mentioned by participants in the suspected-lie condition was behavioral (using the cell notation in Table 1, $S=c/(c+d)$). Then we consider what proportion of all cues mentioned by participants in the discovered-lie condition was behavioral ($D=a/(a+b)$). Next, we need to calculate the difference: $RD=S-D$. Note that RD can range between -1 and 1 , and that if the proportion of behavioral cues is the same in both conditions, then $RD=0$. In Novotny et al.'s (2018) Study 1, $RD=.44$. In Study 2, $RD=.24$. All of those findings show that even though in the suspected-lie condition behavioral cues did not outnumber non-behavioral information, the participants' tendency to mention behavioral cues (relative to non-behavioral information) was much larger in the suspected-lie condition than in the discovered-lie condition.⁴

Novotny et al.'s (2018) Findings in Context: Meta-analytic Thinking

Any single study is just one instance of a population of studies examining the same research question and tracking the same population parameter. Just because of random variability, research outcomes can vary dramatically among studies, even when the true effect is the same. For instance, Cumming (2008; see also Cumming 2012) showed that the sampling variability of p values across replications of the same experiment is extremely large, such that in reality p values give only very vague information. Replication and the progressive accumulation of data permit increasingly more precise estimates; thus, researchers should refrain from making strong claims based on individual studies, focusing instead on the accumulated evidence. This is what Cumming (2012) calls *meta-analytic thinking*: "The consideration of any result in relation to previous results on the same or similar questions, and awareness that combination with future results is likely to be valuable" (p. 9).

To our knowledge, research has never previously examined whether participants report more behavioral cues than non-behavioral information when asked to recall how they suspected a lie. But the kind of information provided when asked to recall a discovered lie has

³ Recall that, as mentioned earlier, in Study 1 Novotny et al. (2018) reported percentages of *lies* discovered from each type of clue (behavioral/non-behavioral), while in Study 2 they reported the percentage of *clues* of each type reported by the participants. Percentages of lies vs. of clues can differ, and therefore can yield slightly different outcomes when calculating OR and RD . This problem prevented us from meta-analytically combining the OR from both studies to have a better estimate of the true effect.

⁴ We are aware that for Novotny et al.'s Study 1 we did not use OR and RD the way it is normally used. To illustrate, medical research focuses on the proportion of participants in the control group who contract (a) and who do not contract (b) a disease, as well as on the proportion of participants in the treatment group who contract (c) and who do not contract (d) the same disease. Note that in the medical research case, as well as in Novotny et al.'s Study 2, $a+b=1$, and $c+d=1$. But this does not happen with Novotny et al.'s Study 1 data, as some lies could be detected from both behavioral and non-behavioral information. Specifically, for the current Study 1 calculations, $a+b=1.04$, and $c+d=1.11$. However, if the proportion of behavioral relative to non-behavioral clues is the same in both conditions, OR is still 1 and RD is 0. Also, the maximum possible range for RD is still -1 to 1 . Finally, the deviations from 1 are minute. Therefore, we do believe that OR and RD are still informative for Study 1.

been explored previously by Park et al. (2002) and Masip and Herrero (2015). Thus, in line with meta-analytic thinking, it might be enlightening to combine Novotny et al.'s (2018) outcomes relative to the latter question with those of prior studies. To do so, we meta-analyzed the proportions of (a) lies detected with each kind of information (behavioral, non-behavioral), and (b) cues of each kind (behavioral, non-behavioral) mentioned by the participants in all the available studies. Caution is warranted concerning the latter analyses because whereas each participant reported on only one lie, each participant could (and many did) mention more than one cue. Therefore, when using cues (rather than lies) as the unit of analysis, the independence of the observations is violated.

We used the logit method to combine the proportions. That is, the observed proportions were converted to logits, all analyses (i.e., calculation of standard errors, inverse variances, weighted mean logits, confidence intervals, and homogeneity analyses) were conducted on the logits, and then the mean logits across studies, as well as the confidence intervals, were transformed back to proportions (see Lipsey and Wilson 2001). We could not include all studies in all meta-analyses because some data were missing (see Table 6). Also, in the study by Masip and Herrero (2015), police officers were asked about two lies: one discovered in personal contexts and one discovered in professional contexts. In order not to duplicate participants, only the personal-context lies (which were of the kind examined in the other studies) were included in the meta-analyses. Note, however, that among Masip and Herrero's officers, the percentages for lies detected in professional situations (over the total number of lies: 38.10% for behavioral cues, 95.24% for non-behavioral information; over the total number of cues: 29.82% for behavioral cues, 61.40% for non-behavioral information) were similar to the percentages for personal-context lies reported in Table 6.⁵

Table 6 shows the outcomes of the original studies as well as the meta-analytical results. The combined percentage of *lies* discovered from behavioral cues across studies ranges between 15.59%, 95% CI [11.62%, 20.60%] and 16.44%, 95% CI [12.12%, 21.91%] (depending on whether we consider Novotny et al.'s 5.33% or 1.33% rate). For non-behavioral information, the meta-analytic percentage is 84.00%, 95% CI [72.34%, 91.33%]. The combined percentage of behavioral *cues* reported by participants as the reason for discovery is 17.07%, 95% CI [13.43%, 21.25%], while the percentage of non-behavioral cues is 81.68%, 95% CI [77.18%, 85.46%]. It is clear that the figures are similar regardless of whether one considers the combined percentage of lies or the combined percentage of cues. Table 6 also shows that all estimates are heterogeneous. Unfortunately, not enough studies are available to identify relevant moderators and conduct moderator analyses.

⁵ As the current article was under review, a paper by Levine and Daiku (2018) was published where the authors examined, among other issues, how customs agents discovered a lie in the past. The answers were classified as “behavioral cues”, “communication content and evidence”, and “interaction/questioning.” While the two first categories are similar to the current “behavioral” and “non-behavioral” categories, several reasons prevented us from including Levine and Daiku's data in the meta-analysis. First, Levine and Daiku's “communication content and evidence” also included “logical consistency” (i.e., absence of verbal contradictions), which is definitely a behavioral cue. Second, in Levine and Daiku's study the same response could be simultaneously coded as pertaining to both the “behavioral” and the “content and evidence” categories. Third, presumably customs agents reported on professional lies; note that we used the personal-context lies (rather than the professional-context lies) of Masip and Herrero's (2015) police officers to keep the meta-analysis uniform in terms of the kind of lies considered. Still, in line with the findings of the current meta-analysis, Levine and Daiku's participants reported more items pertaining to the “communication content and evidence” category (98%) than to the “behavioral cues” category (43%).

Table 6 Individual-study data and meta-analytic integration of studies examining the kind of information (behavioral vs. non-behavioral) mentioned by participants asked to indicate how they discovered a lie

Study	Over the total number of lies				Over the total number of cues			
	Behavioral		Non-behavioral		Behavioral		Non-behavioral	
	N ^a	Percentage	Frequency	Percentage	N ^b	Frequency	Percentage	Percentage
Park et al. (2002)	194	11.34	na	na	266	22	8.27	242
Masip and Herrero (2015)								
Community members	22	40.91	16	72.73	32	15	46.88	16
Police officers ^c	21 ^d	33.33	18	85.71	39	13	33.33	24
Novotny et al. (2018)								
Study 1	75	1.33 to 5.33	74	98.67	na	na	na	na
Study 2	72	na	na	na	93	13	13.98	80
Combined percentages (weighted)		16.44% to 15.59%	84.00%			17.07%		81.68%
95% CI		[12.12%, 21.91%] to [11.62%, 20.60%]	[72.34%, 91.33%]			[13.43%, 21.45%]		[77.18%, 85.46%]
Heterogeneity test		Q = 23.17 to Q = 21.88, p < .001	Q = 8.96, p = .011			Q = 38.08, p < .001		Q = 43.27, p < .001

“na” = data not available

^aTotal number of lies (each participant reported on one lie). ^bTotal number of cues (many participants reported more than one cue). ^cThe current data correspond to officers reporting about a lie discovered in their personal lives. Masip and Herrero also asked officers about a lie discovered during their professional duties. To avoid duplication of participants while keeping the context of the lies similar across studies (personal contexts), we excluded the officers’ professional lies. ^dThe number of police participants in Masip and Herrero was 22; however, one officer left the personal-context-lie questionnaire blank

Summary and Conclusions

Concerns have been raised that laboratory experiments do not reflect the way lies are actually detected in real life (Levine 2018). For example, outside the laboratory human lie detectors have access to information other than behavioral cues, and it is that information (rather than verbal and nonverbal behavior) which people use to successfully detect deception (Levine and Daiku 2018; Masip and Herrero 2015; Novotny et al. 2018; Park et al. 2002). However, Novotny et al. (2018) conceived lie detection as a two-step process wherein behavioral cues elicit suspicion and, subsequently, the individual seeks for harder, non-behavioral evidence to support their suspicion. To test these notions, Novotny et al. conducted two studies in which they asked participants to indicate either how they suspected that someone lied to them or how they detected that someone lied to them. In both studies, Novotny et al. predicted that participants in the discovered-lie condition would mention non-behavioral information more often than behavioral cues (H1a), while participants in the suspected-lie condition would mention behavioral cues more often than non-behavioral information (H1b). They stated they found support for both hypotheses.

In this report, we highlighted some questionable aspects in Novotny et al.'s paper, conducted alternative analyses based on the frequencies and percentages that they reported, and meta-analytically combined their data with those of previous studies. Concerning H1a, the combined evidence from five different samples (involving college students, community members, and police officers) in two different countries indicates that when participants are asked to report how they discovered a lie, they mention behavioral cues only about 16% of the time and non-behavioral information slightly over 80% of the time. Concerning H1b, the only available evidence comes from Novotny et al.'s studies. Analyses of their data failed to support H1b: Apparently, when participants are asked to report how they suspected someone was lying to them, they report behavioral cues as often (Study 1) or presumably less often (Study 2) than non-behavioral information. This is at odds with Novotny et al.'s claim that in the suspected-lie condition the participants mentioned more behavioral cues than non-behavioral evidence.

Still, Novotny et al. are correct in suggesting that behavioral cues may play a greater role in eliciting suspicion than confirmation. The likelihood of mentioning behavioral cues (relative to non-behavioral information) was indeed substantially greater in the suspected-lie than in the discovered-lie condition, even though behavioral cues prevailed in both conditions. We believe this is an important finding that helps clarify the processes involved in lie detection in real life—that is, outside the artificial context of the laboratory.

Acknowledgements The authors are grateful to Dr. Iris Blandón-Gitlin (California State University, Fullerton, USA) and Ms. Ana I. Jiménez (University of Salamanca, Spain) for their comments on a previous draft of this manuscript.

References

- Bond, C. F., Jr., & DePaulo, B. M. (2006). Accuracy of deception judgments. *Personality and Social Psychology Review*, *10*, 214–234. https://doi.org/10.1207/s15327957pspr1003_2.
- Cumming, G. (2008). Replication and *p* intervals: *p* values predict the future only vaguely, but confidence intervals do much better. *Perspectives on Psychological Science*, *3*, 286–300. <https://doi.org/10.1111/j.1745-6924.2008.00079.x>.

- Cumming, G. (2012). *Understanding the new statistics. Effect sizes, confidence intervals, and meta-analysis*. New York, NY: Routledge.
- DePaulo, B. M., Lindsay, J. J., Malone, B. E., Muhlenbruck, L., Charlton, K., & Cooper, H. (2003). Cues to deception. *Psychological Bulletin*, *129*, 74–118. <https://doi.org/10.1037/0033-2909.129.1.74>.
- Ferguson, C. J. (2009). An effect size primer: A guide for clinicians and researchers. *Professional Psychology: Research and Practice*, *40*, 532–538. <https://doi.org/10.1037/a0015808>.
- Levine, T. R. (2018). Ecological validity and deception detection research design. *Communication Methods and Measures*, *12*, 45–54. <https://doi.org/10.1080/19312458.2017.1411471>.
- Levine, T. R., & Daiku, Y. (2018). How custom agents really detect lies. *Communication Research Reports*, *36*, 84–92. <https://doi.org/10.1080/08824096.2018.1555523>.
- Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Thousand Oaks, CA: Sage.
- Masip, J., & Herrero, C. (2015). Police detection of deception: Beliefs about behavioral cues to deception are strong even though contextual evidence is more useful. *Journal of Communication*, *65*, 125–145. <https://doi.org/10.1111/jcom.12135>.
- Newcombe, R. G. (1998). Interval estimation for the difference between independent proportions: Comparison of eleven methods. *Statistics in Medicine*, *17*, 873–890. [https://doi.org/10.1002/\(SICI\)1097-0258\(19980430\)17:8<873:AID-SIM779>3.0.CO;2-I](https://doi.org/10.1002/(SICI)1097-0258(19980430)17:8<873:AID-SIM779>3.0.CO;2-I).
- Novotny, E., Carr, Z., Frank, M. G., Dietrich, S. B., Shaddock, T., Cardwell, M., et al. (2018). How people really suspect and discover lies. *Journal of Nonverbal Behavior*, *42*, 41–52. <https://doi.org/10.1007/s10919-017-0263-2>.
- Park, H. S., Levine, T. R., McCornack, S. A., Morrison, K., & Ferrara, S. (2002). How people really detect lies. *Communication Monographs*, *69*, 144–157. <https://doi.org/10.1080/714041710>.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.