Psicothema

Psicothema 2020, Vol. 32, No. 3, 329-336
doi: 10.7334/psicothema2019.405

ISSN 0214 - 9915 CODEN PSOTEG

www.psicothema.com

Do people detect deception the way they think they do? Replication and extensions

Nuria Sánchez and Jaume Masip
University of Salamanca

Abstract

Background: Research shows that people believe deception can be detected from behavioral cues despite their past experience of detecting lies from non-behavioral, contextual information (evidence, third-person reports, etc.). However, in previous research, the question about beliefs was necessarily general, while the question about revealing information was always about a specific lie. In this study, we addressed this problem. **Method:** Participants first indicated how they believed lies can be detected (beliefs; Questionnaire 1 or Q1). Next, they described either how they, in their past, detected a *specific lie, several lies*, or how they, *in general*, detect lies in their everyday lives (revealing information; Q2). **Results:** Regardless of the focus of Q2, and in line with prior research, behavioral cues were reported less often, and contextual indicators more often, in responding to Q2 than in responding to Q1. However, contrary to prior findings, behavioral cues still predominated in the responses to Q2. **Conclusions:** We found no evidence that the specific-vs.-general focus of the questions changed the pattern of results, which apparently depended solely on whether participants reported beliefs or revealing information. We provide explanations for the prevalence of behavioral cues in Q2 responses, and make suggestions for future research.

Keywords: Deception detection; deception cues; beliefs; contextual information; everyday life.

Resumen

¿La gente detecta las mentiras tal como cree que lo hace? Replicación y ampliaciones. **Antecedentes:** la investigación muestra que las personas creen que la mentira se detecta a partir de claves conductuales pese a haber detectado mentiras en el pasado a partir de información contextual (evidencias, información de terceros...). En dicha investigación previa, la pregunta sobre creencias ha sido general, mientras que la referente a información reveladora ha sido sobre una mentira concreta. Este estudio resuelve este problema. **Método:** los participantes indicaron cómo creían que se pueden detectar mentiras (creencias; Cuestionario 1 o C1). Luego describieron cómo, en el pasado, habían descubierto *una mentira, varias mentiras,* o cómo, *en general*, suelen detectar mentiras en su vida cotidiana (información reveladora; C2). **Resultados:** independientemente de la modalidad de C2, y en línea con la investigación previa, las claves conductuales se mencionaron menos, y los indicadores contextuales más, al responder a C2 que a C1. Sin embargo, se mencionaron más indicios conductuales que contextuales incluso en C2. **Conclusiones:** no hallamos evidencia de que el foco específico o general de las preguntas cambiara el patrón de resultados, que al parecer dependió solo de si se mencionaban creencias o información reveladora. Ofrecemos explicaciones para la prevalencia de claves conductuales en C2 y hacemos sugerencias para la investigación futura.

Palabras clave: detección de mentira; claves del engaño; creencias; información contextual; vida cotidiana.

For decades, scientists have conducted hundreds of laboratory experiments examining whether deception can be detected from behavioral indicators (for reviews, see Masip, 2017; Nortje & Tredoux, 2019; Vrij, 2008). Two main conclusions of this research are that behavioral indicators are barely related to truthfulness (DePaulo et al., 2003; Sporer & Schwandt, 2006, 2007), and that humans are poor lie detectors. Indeed, a major meta-analysis (Bond & DePaulo, 2006) revealed that veracity judgments were correct just 54% of the time (compared to 50% chance accuracy and 100% perfect accuracy).

Received: December 25, 2019 • Accepted: March 3, 2020
Corresponding author: Jaume Masip
Facultad de Psicología
University of Salamanca
37005 Salamanca (Spain)
e-mail: jmasip@usal.es

Yet, it has been argued that laboratory experiments fail to mirror real-life lie-detection circumstances (e.g., Levine, 2018). In the laboratory, observers are requested to judge the veracity of video-recorded, audio-recorded, or written statements of unacquainted senders, and to do so immediately. Under these circumstances, the only information available to observers is the senders' verbal and nonverbal behavior. Conversely, in real life, deception targets might know the senders and do not need to make an immediate judgment. They can question the senders, learn about their possible motives to deceive, ask informants, and can otherwise search for evidence confirming or refuting the senders' statements. In everyday situations, present a suspicion of deceit (as is typically the case in laboratory experiments because participants are explicitly asked to judge veracity), access to information other than fallible behavioral cues can increase accuracy.

Indeed, in a seminal study, Park, Levine, McCornack, Morrison, and Ferrara (2002) asked students to recall a lie they had detected

in the past and to indicate how they detected it. Results showed that, typically, in everyday life, lies are discovered long after being told, and are not detected from behavioral cues but from non-behavioral, contextual information, such as third-person reports, tangible evidence, the liars' confession, or inconsistencies between the lie and the detector's knowledge. Contextual information either reveals deception or can be used to compare the senders' messages to assess their veracity (see Blair, Levine, Reimer, & McCluskey, 2012). This latter strategy has been called *content in context* by Blair, Levine, and Shaw (2010; see also Levine, 2020), and is analogous to Stiff et al.'s (1989) *situational familiarity hypothesis*, which states that contextual knowledge allows lie detectors to "visualize" the situation and then judge the plausibility of the message (for empirical tests, see Reinhard, Sporer, Scharmach, & Marksteiner, 2011). Indeed, there is evidence that access to contextual information actually increases the accuracy of veracity judgments relative to behavioral cues (Blair et al., 2010; Blair, Reimer, & Levine, 2018; Bond, Howard, Hutchison, & Masip, 2013).

The preponderance of contextual information in signaling deception in real-life contexts first reported by Park et al. (2002) has been confirmed by subsequent research (Masip & Herrero, 2015; Novotny et al., 2018; for a meta-analytical integration, see Masip & Sánchez, 2019). Both this preponderance and meta-analytical findings showing that behavior is barely related to veracity (DePaulo et al., 2003; Sporer & Schwandt, 2006, 2007; see also Luke, 2019) are at odds with the strong popular belief that honesty is revealed through behavioral cues (e.g., Global Deception Research Team, 2006). Before asking respondents to indicate how they had detected a lie in the past (revealing information), Masip and Herrero (2015) asked them the open question how they believed lies can be detected (beliefs). While revealing information was mostly contextual (rather than behavioral), believed deception cues were mostly behavioral. In other words, the allure of behavioral cues as indicators of deception is so strong that it is immune to the individual's personal experience that everyday lies are detected mostly from contextual information rather than from behavioral cues.

A limitation of Masip and Herrero's (2015) study is that while the question about beliefs ("Please indicate how you believe lies can be detected") was general, the question about revealing information was about a specific lie. This raises the question whether differences are a result of asking about beliefs versus revealing information, or a result of asking a general, abstract question versus a question focused on a specific lie. In the current study, we addressed this issue. We used the same procedure as Masip and Herrero (2015), except that we manipulated the focus of the revealing-information question. Specifically, we asked participants to report either (a) how they detected a *specific lie*, (b) how they detected *several lies*, or (c) how they, *in general*, detect lies in everyday life. The specific-lie question was the same used by Masip and Herrero; therefore, we expected to replicate their findings. For the several-lies condition, we also expected to replicate Masip and Herrero's outcomes, thus showing that these outcomes were not caused by the specific versus general focus of the question, but by whether participants reported revealing information or beliefs. Finally, for the "general" condition, the prediction was less clear; while the question was indeed further removed from a specific deception instance, it considerably overlapped with the question about beliefs.

An additional goal of this research was to examine whether Masip and Herrero's (2015) findings (obtained with middle-aged, mostly [77%] male, community members and police officers) could be replicated with a different sample of participants (younger, mostly female, college students).

Method

Participants

Seventy college students (70% females; M_{age} = 19.60, SD = 2.00) volunteered to participate in exchange for an academic incentive. They were criminology (66%) and psychology (34%) students. None of them had still taken the specific courses where lie detection is covered.

Instruments

Questionnaire 1 (Q1) was to collect information about beliefs. After some short demographic questions, it contained the open prompt "Please indicate how you believe lies can be detected." Questionnaire 2 (Q2) focused on revealing information. It had three versions. The *specific-lie-condition* version asked participants to think of a lie they had detected in the past and then, with this in mind, to describe in detail how they detected the lie. The *several-lies-condition* version asked participants to think of several situations in which they had detected a lie and then, with this in mind, to describe in detail how they detected these lies. The *general-condition* version asked participants to think about how, in general, they discover lies in their everyday lives; it explicitly asked participants not to think of a specific instance of deception, but in general terms. Then, with this in mind, participants had to describe in detail how they normally detect lies. Both Q1 and Q2 had a box where each individual participant wrote a word or number of their choice; this was to allow us to put together the two questionnaires of each individual while preserving the participants' anonymity.

Procedure

Data collection. The study was conducted in accordance with national and international ethical regulations. The data were collected in the context of a practical lecture. The students sat apart from each other and were invited to participate. After signing an informed-consent form, they were instructed not to talk to their peers until the end of the session, and were given Q1. After they all had finished, Q1 was collected and Q2 was handed out. The three versions of Q2 were alternated. After all the participants had finished, we collected Q2, thanked the participants and debriefed them.

Coding. Two naïve research assistants were instructed to code the participants' answers using a slightly modified version of Masip and Herrero's (2015) coding scheme. The coding scheme differentiated between several kinds of *behavioral cues* (visible, verbal, paralinguistic, physiological, and unspecified-behavioral), *contextual information* (third-person information, evidence, confession, inconsistency with knowledge, and other-contextual), and the residual *other information* category. We added an additional contextual category—*dispositional honesty/dishonesty*—and slightly improved the definitions of some other categories.

Disagreements between the coders were resolved by discussion. The coding categories and reliabilities are displayed in Table 1. Reliabilities were calculated considering both dichotomous data (whether [1] or not [0] a participant mentioned one or more indicators pertaining to a specific category) and frequencies (the number of indicators pertaining to a specific category mentioned by the participant). Of note, Masip and Herrero (2015) included *physiological cues* (blushing, trembling…) within the *behavioral-cues* higher-level category (see Table 1), but after the coding for the current study was complete, we noticed that, unlike in Masip and Herrero, a number of respondents had mentioned psychophysiological measurements (e.g., the polygraph). These responses had been assigned by the coders to the *physiological-cues* category, but, unlike visible physiological cues, they cannot be considered observable behavioral cues—equipment is needed to measure them. Therefore, we asked two additional naïve coders to re-code responses first placed in the *physiological* category as either *visible cues* (e.g., blushing, trembling…) or as *physiological measurements* (e.g., polygraph examination). The latter category was considered as a separate higher-level category besides *behavioral cues* and *contextual information* (see Table 2).

Table 1 shows that reliability was very high, except for the residual category *other-contextual* (which had few cases), and for *inconsistency with knowledge*. For the two higher-level categories of interest (*behavioral cues* and *contextual information*) reliability was also very high. Reliability was also good for the re-coding of *physiological cues* as either *visible* (*Kappa* = .87, percent agreement = 93.61, *ICC (2,2)* = .94, *r* = .89) or *physiological measurements* (*Kappa* = .91, percent agreement = 95.74, *ICC (2,2)* = .96, *r* = .92).

Table 1
Inter-rater Reliability

Type of Information	Dichotomous data		Frequency data	
	Kappa	Percent agreement	ICC (2,2)	r
Behavioral Cues	*.94*	*99.28*	*.99*	*.99*
Visible	.96	98.57	.99	.97
Verbal	.91	95.71	.96	.92
Paralinguistic	.92	96.43	.96	.93
Physiological	.92	96.43	.98	.96
Unspecified-behavioral	.86	92.86	.96	.92
Contextual Information	*.81*	*90.72*	*.96*	*.93*
Third-person information	.89	97.86	.96	.92
Evidence	.72	90.72	.85	.76
Confession	.97	99.28	.98	.97
Inconsistency with knowledge	.65	91.43	.58	.41
Dispositional (dis)honesty	1.00	100.00	1.00	1.00
Other-contextual	-.02	95.71	-.04	-.02
Other Information	*.59*	*97.14*	*.75*	*.65*

Data analysis

We conducted a Condition (one lie vs. several lies vs. general) × Questionnaire (Q1-beliefs vs. Q2-revealing information) × Type of Information (behavioral vs. contextual) test, with repeated measures in the latter two variables, on the *frequencies* for behavioral vs. contextual information. We also ran an analogous test on the *dichotomous* data. In line with Masip and Herrero (2015), we expected the Questionnaire × Type of Information interaction to be significant. Also, if this interaction occurred independently of the focus (on a specific lie vs. more general) of the question used in Q2, the three-way interaction would not be significant. Because the distributions seriously violated the assumption of normality, we refrained from using an analysis of variance (ANOVA). Instead, we conducted an ANOVA-type test with Noguchi, Gel, Brunner, and Konietschke's (2012) nparLD package for R.

Noguchi et al.'s (2012) package uses a nonparametric, rank-based method to conduct robust, distribution-free tests for main effects and interactions in repeated-measure and mixed designs. These methods were first developed by Brunner and Puri (2001), and by Brunner, Domhof, and Langer (2002), and were incorporated into a SAS/IML macro library. nparLD, is an R version of that library (Noguchi et al., 2012). It calculates an ANOVA-type statistic (*ATS*) that tests the null hypothesis that the groups being compared have identical distributions and the same relative treatment effects (*RTE*s) (Erceg-Hurn & Mirosevich, 2008). *RTE*s can range between 0 and 1, and reflect "the tendency for participants in one group to have higher (or lower) scores on the dependent variable, compared with the scores of all participants in a study" (Erceg-Hurn & Mirosevich, 2008, p. 597). In other words, the higher the *RTE*, the higher the probability that a randomly chosen observation from the whole dataset has a smaller value than a randomly chosen observation from the condition with that *RTE* (Noguchi et al., 2012). If the null hypothesis is true, then all conditions should have *RTE* = .50; *RTEs* < .05 denote relatively low values in the dependent variable, while *RTEs* > .05 denote relatively high values.

Pairwise comparisons to decompose significant interactions were performed with Rogmann's (2013) orddom R package, which calculates Cliff's (1993, 1996) delta and the corresponding Cohen's (1988) *d*. Cliffs delta (*d*) is a statistic that "compares the number of times a score from one group or condition is higher than one from the other, compared with the reverse" (Cliff, 1993, p. 494). It ranges from -1 to +1, and requires no assumptions.

Results

The numbers of participants originally allocated to the specific-lie, the several-lies, and the general condition were 23, 24, and 23, respectively. However, it became apparent in reading the respondents' answers that one specific-lie participant described several lies, and that several participants in the several-lies condition either described just one lie (n = 4) or explained how to detect deception in general (n = 8). After moving these participants to the corresponding condition, sample sizes were 26, 13, and 31 for the specific-lie, the several-lies, and the general conditions respectively. Neither gender, χ^2 (2) = 1.04, p = .594, *phi* = .122, nor age, F (2, 67) = 0.98, p = .380, η_p^2 = .028, nor academic background (criminology vs. psychology), χ^2 (2) = 0.68, p = .711, *phi* = .099, differed significantly across conditions.

Table 2 displays the mean and median frequencies for each kind of information, as well as the percentage of participants in each condition mentioning each specific kind of information. It is apparent that the most cited kind of indicator was *visible cues*, followed by *verbal cues*. Among the contextual information, *evidence* was mentioned most often in all cases except in the specific-lie condition-Q1 and the general condition-Q2, where

third-person information and inconsistency with knowledge were, respectively, mentioned most often (Table 2).

Frequencies

The ANOVA-type test on frequencies revealed a significant main effect for questionnaire, *ATS (1)* = 9.92, p = .002, indicative that fewer indicators were mentioned in responding to Q1 (*RTE* = .48) than in responding to Q2 (*RTE* = .52). The main effect for information type was also significant, *ATS (1)* = 129.31, p < .001; more behavioral cues were mentioned (*RTE* = .70) than contextual indicators (*RTE* = .30). More interestingly, as predicted, the Questionnaire × Information-Type interaction was significant (Figure 1a), *ATS (1)* = 33.50, p < .001, while the three-way interaction was not, *ATS (1)* = 2.68, p = .078. This latter outcome suggests that the two-way interaction held regardless of condition.

Table 2
Descriptive Statistics

Type of Information	Specific-lie condition (n = 26)						Several-lies condition (n = 13)						General condition (n = 31)					
	Q1			Q2			Q1			Q2			Q1			Q2		
	M	Mdn	%	M	Mdn	%	M	Mdn	%	M	Mdn	%	M	Mdn	%	M	Mdn	%
Behavioral Cues	***4.81***	***4.50***	***100***	***3.08***	***3.00***	***81***	***4.69***	***5.00***	***100***	***4.38***	***3.00***	***92***	***4.42***	***4.00***	***100***	***4.48***	***4.00***	***94***
Visible	2.38	2.00	92	0.96	0.50	50	2.31	3.00	100	1.54	1.00	85	2.35	2.00	97	1.61	1.00	71
Verbal	1.08	1.00	62	0.96	1.00	54	0.92	1.00	62	1.54	1.00	62	0.71	1.00	55	1.32	1.00	77
Paralinguistic	0.42	0.00	31	0.46	0.00	31	0.38	0.00	31	0.54	0.00	38	0.42	0.00	32	0.58	0.00	45
Unspecified-behavioral	0.92	1.00	58	0.69	0.50	50	1.08	1.00	77	0.77	0.00	31	0.94	1.00	58	0.97	1.00	55
Contextual Information	***0.31***	***0.00***	***19***	***1.81***	***2.00***	***73***	***0.31***	***0.00***	***15***	***1.85***	***1.00***	***69***	***0.13***	***0.00***	***13***	***0.90***	***1.00***	***55***
Third-person information	0.12	0.00	12	0.23	0.00	23	0.08	0.00	8	0.38	0.00	23	0.00	0.00	0	0.16	0.00	13
Evidence	0.04	0.00	4	0.77	0.00	46	0.15	0.00	15	0.77	0.00	38	0.13	0.00	13	0.26	0.00	23
Confession	0.04	0.00	4	0.42	0.00	42	0.08	0.00	8	0.23	0.00	23	0.00	0.00	0	0.03	0.00	3
Inconsistency with knowledge	0.04	0.00	4	0.27	0.00	23	0.00	0.00	0	0.31	0.00	31	0.00	0.00	0	0.32	0.00	26
Dispositional (dis)honesty	0.04	0.00	4	0.00	0.00	0	0.00	0.00	0	0.08	0.00	8	0.00	0.00	0	0.03	0.00	3
Other-contextual	0.04	0.00	4	0.12	0.00	8	0.00	0.00	0	0.08	0.00	8	0.00	0.00	0	0.10	0.00	10
Physiological Measurements	***0.62***	***0.00***	***42***	***0.00***	***0.00***	***0***	***0.38***	***0.00***	***31***	***0.00***	***0.00***	***0***	***0.45***	***0.00***	***42***	***0.00***	***0.00***	***0***
Other Information	***0.12***	***0.00***	***8***	***0.00***	***0.00***	***0***	***0.00***	***0.00***	***0***	***0.00***	***0.00***	***0***	***0.06***	***0.00***	***6***	***0.03***	***0.00***	***3***
TOTAL	5.85	5.00	100	4.88	4.50	100	5.38	6.00	100	6.23	5.00	100	5.06	5.00	100	5.42	5.00	100

Note: M: Mean number of indicators; *Mdn*: Median number of indicators; *%*: Percentage of participants mentioning indicators of the specific kind

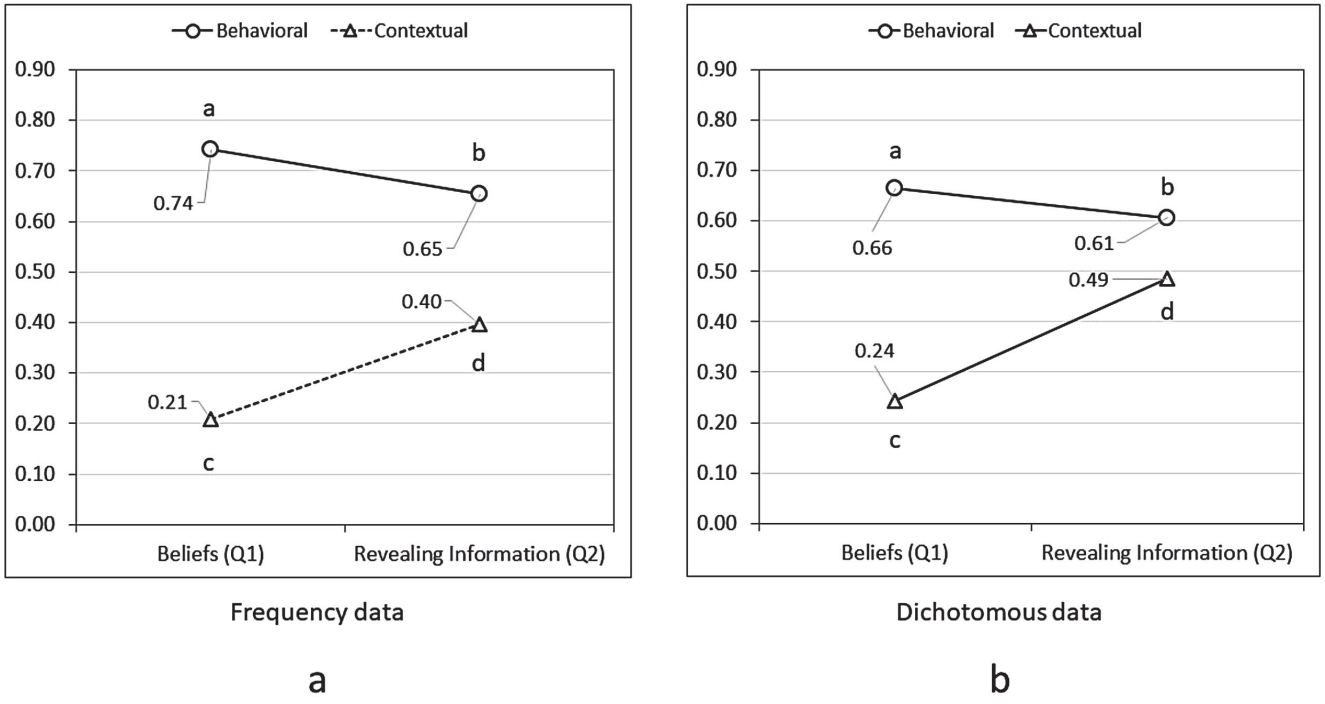

Figure 1. Relative treatment effects (RTEs) for the Questionnaire x Type of Information interaction

It is unlikely that the non-significance of this interaction was caused by insufficient power. Sample size was not remarkably small and nonparametric tests are generally much more powerful than parametric tests (e.g., Brunner et al., 2002; Erceg-Hurn & Mirosevich, 2008). In addition, Noguchi et al.'s (2012) ANOVA-type test "maintains an accurate size of the test even for small sample sizes ($n \geq 7$)" (Noguchi et al., 2012, p. 8).

Table 3 displays the results of the pairwise comparisons decomposing the significant two-way interaction. All comparisons were significant, with more behavioral cues being mentioned than contextual indicators when the participants were asked about both, beliefs (with very large effect sizes) and revealing information (with smaller, but still rather large effect sizes). Also, when asked about revealing information (Q2), participants mentioned significantly fewer behavioral cues and significantly more contextual indicators than when they were asked about beliefs (Q1).

Dichotomous data

The analyses on the dichotomous data yielded similar results. Again, the ANOVA-type test revealed significant main effects for questionnaire, *ATS (1)* = 25.36, $p < .001$ (for Q1, *RTE* = .45; for Q2, *RTE* = .55), and for information type, *ATS (1)* = 98.55, $p < .001$ (for behavioral cues, *RTE* = .64; for contextual information, *RTE* = .36), and a significant Questionnaire x Information Type interaction, *ATS (1)* = 63.76, $p < .001$ (Figure 1b). The outcomes of pairwise-comparison tests for this interaction are shown in Table 4. The three-way interaction was again not significant, *ATS (1)* = 0.87, $p = .416$.

Discussion

Masip and Herrero (2015) found that more respondents mentioned behavioral cues, and fewer mentioned contextual information, in responding to Q1 (beliefs) than in responding to Q2 (revealing information). However, in their study, the Q1 question was general, while the Q2 question focused on a specific lie. Therefore, there was a confound between asking about beliefs vs. revealing information and the general vs. specific focus of the questions. We addressed this issue by using three versions of the revealing information question; one that asked about a specific lie, one that asked about several lies, and one that was nearly as general as the Q1 question. As expected, our results revealed a decrease in both the frequency of behavioral cues and the percentage of participants mentioning them in responding to Q2 compared to Q1, as well as parallel increases for contextual cues (Figure 1). Importantly, we found no evidence that these effects depended on the focus of the Q2 question. Thus, it is unlikely that Masip and Herrero's (2015) findings were caused by a difference between the focus of the belief question (general) and that of the revealing-indicators question (specific).

We also examined whether Masip and Herrero's (2015) outcomes were replicated with participants from a different population. As expected, the superiority of behavioral cues over contextual information when the participants were asked about their general beliefs (Q1) was replicated. This outcome can be due to behavioral (particularly nonverbal) cues being more global in their use (i.e., independent from a specific context) than contextual indicators, as well as easier to process than other kinds of information (see Reinhard & Sporer, 2008, 2010). It can also be a result of worldwide socialization practices fostering in children the belief that lying elicits negative emotions that are revealed through visible behavior (see Global Deception Research Team, 2006).

As mentioned above, we also replicated Masip and Herrero's (2015) finding of a decrease for behavioral cues, and an increase for contextual information, in Q2 (relative to Q1). However, unlike Masip and Herrero, we found no evidence for the superiority of contextual information over behavioral cues in responding to

Table 3
Pairwise Comparisons for the Questionnaire × Information-Type Interaction with the Frequency Data

Comparison	Cliff's d (95% CI)	SD of Cliff's d	Z score	p (two tailed)	Cohen's d (95% CI)
a > *b*					
Within	.24 (.02, .46)	.11	2.20	.031	0.35 (0.03, 0.77)
Between	.17 (.003, .35)	.09	2.03	.046	0.24 (0.004, 0.53)
Combined	0.42 (0.04, 0.80)	0.19	2.20	.031	–
c < *d*					
Within	-.59 (-.71, -.46)	.06	-9.34	< .001	-0.58 (-0.67, -0.47)
Between	-.51 (-.63, -.39)	.06	-8.58	< .001	-0.52 (-0.61, -0.41)
Combined	-1.10 (-1.33, -0.86)	0.12	-9.34	< .001	–
a > *c*					
Within	1.00 (.89, 1.00)	.00	∞	< .001	∞ (2.60, ∞)
Between	.98 (.97, .98)	.01	123.13	< .001	4.21 (3.68, 5.73)
Combined	1.98 (1.97, 2.00)	0.01	248.52	< .001	–
b > *d*					
Within	.39 (.17, .60)	.11	3.56	< .001	0.61 (0.23, 1.14)
Between	.57 (.39, .75)	.09	6.34	< .001	1.05 (0.62, 1.69)
Combined	0.96 (0.57, 1.35)	0.19	4.91	< .001	–

Note: As shown in Figure 1, *a*: Beliefs – Behavioral; *b*: Revealing information – Behavioral; *c*: Beliefs – Contextual; and *d*: Revealing information – Contextual. Within: For repeated-measure designs, Cliff's d_w is the "difference between the proportion of individual subjects who change in one direction and the proportion of individuals who change in the other" (Cliff, 1996, p. 159). Between: For repeated-measure designs, Cliff's d_b is "the extent to which the overall distribution has moved, except for the self-comparisons" (Cliff, 1996, p. 160)

Table 4
Pairwise Comparisons for the Questionnaire × Information-Type Interaction with the Dichotomous Data

Comparison	Cliff's d (95% CI)	SD of Cliff's d	Z score	p (two tailed)	Cohen's d (95% CI)
a > *b*					
Within	.11 (.04, .19)	.04	2.98	.004	0.15 (0.05, 0.26)
Between	.11 (.04, .19)	.04	2.98	.004	0.15 (0.05, 0.26)
Combined	0.23 (0.08, 0.38)	0.08	2.98	.004	–
c < *d*					
Within	-.49 (-.61, -.37)	.06	-8.07	< .001	-0.49 (-0.59, -0.39)
Between	-.49 (-.61, -.37)	.06	-8.07	< .001	-0.49 (-0.59, -0.39)
Combined	-0.97 (-1.21, -0.73)	0.12	-8.07	< .001	–
a > *c*					
Within	.84 (.76, .93)	.04	19.24	< .001	2.20 (1.71, 3.03)
Between	.84 (.76, .93)	.04	19.24	< .001	2.20 (1.71, 3.03)
Combined	1.69 (1.51, 1.86)	0.09	19.24	< .001	–
b > *d*					
Within	.24 (.09, .40)	.08	3.14	.002	0.35 (0.12, 0.63)
Between	.24 (.09, .40)	.08	3.14	.002	0.35 (0.12, 0.63)
Combined	0.49 (0.18, 0.79)	0.15	3.14	.002	–

Note: As shown in Figure 1, *a*: Beliefs – Behavioral; *b*: Revealing information – Behavioral; *c*: Beliefs – Contextual; and *d*: Revealing information – Contextual. Within: For repeated-measure designs, Cliff's d_w is the "difference between the proportion of individual subjects who change in one direction and the proportion of individuals who change in the other" (Cliff, 1996, p. 159). Between: For repeated-measure designs, Cliff's d_b is "the extent to which the overall distribution has moved, except for the self-comparisons" (Cliff, 1996, p. 160)

Q2. On the contrary, we found a significant difference in favor of behavioral cues (though this difference was considerably smaller than for Q1). This finding is remarkable because all prior research has found that, when participants are asked how they detected lies in the past (revealing information), contextual indicators are mentioned more frequently, or by more participants, than behavioral cues (Masip & Herrero, 2015; Novotny et al., 2018; Park et al., 2002; for an overview, see Masip & Sánchez, 2019).

There are two factors which, in combination, might have been responsible for this unexpected finding. First, because most participants were criminology students, they might have been particularly conscientious in reporting all kinds of deception cues, which might have led them to report not only detection indicators, but also mere suspicion ones. This is apparent in reading several respondents' answers. Consider this example (Participant #007; the surface level details have been edited to maintain the individuals' anonymity, but indicators and strategies have been left untouched):

> *There was that friend who did not want to tell me about her visit to the doctor. She told me she did not recall or told me about it in a general way. It was like that several times, which was surprising to me because in the short term you always remember what your doctor told you. One day I decided to ask her sister, who told me that my friend was undergoing a number of diagnostic tests because they suspected she could be suffering from breast cancer. That's why she refused to give direct answers or switched topic every time I asked her. She kept doing this for a while, but I was more alert and noticed it, until one day I confronted her and she had no option but to acknowledge that she had been hiding that information from me.*

Clearly, this response contains several *suspicion cues* (evasive or vague responses given by the friend, the friend switching topic…), two *strategies* to corroborate the suspicion (asking the friend's sister and confronting the friend), and two *detection indicators* (the sister's information and the friend's ultimate confession). There were several replies similar to this one, as well as many where it was uncertain to us whether most indicators where suspicion or detection ones. Research has shown that more behavioral cues are reported when participants are asked about suspicion indicators than when they are asked about detection indicators (Novotny et al., 2018; Masip & Sánchez, 2019); therefore, the inclusion of numerous indicators of mere suspicion by our participants might have increased the number of behavioral cues in Q2.

This issue underscores the need to conceptually distinguish between suspicion indicators, the detector's strategies (to corroborate their suspicion), and detection indicators. Also, researchers should explicitly tell participants about the distinction between suspicion and detection indicators, so that the participants fully understand what specific kind of indicators researchers are asking for.

The second reason why, overall, we found that the Q2 responses contained more behavioral than contextual indicators lies in the general condition's responses. As noted, the general pattern of results was the same across conditions, but the separate effects might have been somewhat stronger in certain conditions than others. Specifically, visual inspection of Table 2 suggests that, for the general condition, the decrease for behavioral cues in Q2 (relative to Q1) was meager. Also, the increase for contextual information, though substantial, looks more modest than for the other conditions (formal calculations of Cliff's *d*s and the associated significance levels supported these impressions; these analyses are available from the corresponding author on request). Thus, even though the Questionnaire × Type of Information interaction was still significant considering the general condition only (for frequency data, $ATS\ (1) = 5.70$, $p = .017$; for dichotomous data, $ATS\ (1) = 18.54$, $p < .001$), the trend for this condition to contain some more behavioral cues and fewer contextual indicators than the other conditions might have contributed to the significant difference in Q2 in favor of behavioral cues. These trends for the

general condition might be due to the resemblance between the Q2 question for this condition and the Q1 question.

Readers may think of a third possible reason why behavioral cues were still prevalent over contextual information in the responses to Q2: Q1 might have primed participants to think of general indicators; in replying to Q2, they might have provided the same (or very similar) answers. However, this is unlikely; it should be noted that Masip and Herrero (2015) also asked participants to report their beliefs first, yet they found the usual prevalence of contextual over behavioral cues in responding to Q2.

There is an additional interesting difference between the outcomes of this study and those of Masip and Herrero (2015). A substantial proportion of the current participants mentioned psychophysiological devices and measures in explaining how they believed lies can be detected (Q1). Again, this effect might have emerged because most participants were young criminology students rather than middle-aged ordinary citizens or seasoned local police officers. Although none of our participants had taken any of the courses in their degrees covering lie detection, they presumably had interest in criminalistic issues and watched fictional TV series displaying the polygraph and more sophisticated "high-tech" devices to detect deception.

To conclude, we compared people's beliefs about deception cues (Q1) with revealing information (Q2). We found a large preponderance of behavioral cues when participants reported beliefs. When asked about revealing deception indicators, the participants mentioned fewer behavioral cues and more contextual information than when asked about beliefs. However, behavioral cues were also somewhat predominant among the reported revealing indicators. This latter finding appears to be a consequence of our participants having mentioned suspicion cues in addition to detection indicators, coupled with a tendency for the general condition group to show somewhat more modest changes (compared to the other conditions) in Q2 relative to Q1. In any case, we found no evidence that the specific-vs.-more-general focus of the revealing information question had any substantial effect on the general pattern of results (i.e., on the meaningful Questionnaire × Information Type significant interaction), which replicated across conditions. The study also stresses the need to separate between deception indicators and lie-detection strategies, as well as to distinguish between suspicion and deception indicators. It also underscores the need to be clear about the specific kind of indicator requested from participants. Finally, this study shows that different populations can have different views about how lies can be detected.

Acknowledgements

Thanks are due to Borja Martí, Sheila Rodríguez, Mireia Sánchez and David Sánchez for their assistance; to Rodrigo Carcedo and Emiliano Díez for their suggestions concerning data analysis; to Ana Isabel Jiménez for her comments on the manuscript, and to Fundación Universitaria Behavior & Law for its endorsement of the research proposal.

This research was supported by the Consejería de Educación, Junta de Castilla y León (Spain), Subvenciones Destinadas al Apoyo de los Grupos de Investigación Reconocidos de Universidades Públicas de Castilla y León (Grant Number SA041G19).

References

Blair, J. P., Levine, T. R., & Shaw, A. S. (2010). Content in context improves deception detection accuracy. *Human Communication Research, 36*, 423-442. https://doi.org/10.1111/j.1468-2958.2010.01382.x

Blair, J. P., Levine, T. R., Reimer, T. O., & McCluskey, J. D. (2012). The gap between reality and research. Another look at detecting deception in field settings. *Policing: An International Journal, 35*, 723-740. https://doi.org/10.1108/13639511211275553

Blair, J. P., Reimer, T. O., & Levine, T. R. (2018). The role of consistency in detecting deception: The superiority of correspondence over coherence. *Communication Studies, 69*, 483-498. https://doi.org/10.1080/10510974.2018.1447492

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

Bond, C. F., Jr., Howard, A. R., Hutchison, J. L., & Masip, J. (2013). Overlooking the obvious: Incentives to lie. *Basic and Applied Social Psychology, 35*, 212-221. https://doi.org/10.1080/01973533.2013.764302

Brunner, E., Domhof, S., & Langer, F. (2002). *Nonparametric analysis of longitudinal data in factorial experiments*. New York: Wiley.

Brunner, E., & Puri, M. L. (2001). Nonparametric methods in factorial designs. *Statistical Papers, 42*, 1-52. https://doi.org/10.1007/s003620000039

Cliff, N. (1993). Dominance statistics: Ordinal analyses to answer ordinal questions. *Psychological Bulletin, 114*, 494-509. https://doi.org/10.1037/0033-2909.114.3.494

Cliff, N. (1996). *Ordinal methods for behavioral data analysis*. Mahwah, NJ: Lawrence Erlbaum.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. New York, NY: Academic Press.

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

Erceg-Hurn, D. M., & Mirosevich, V. M. (2008). Modern robust statistical methods. An easy way to maximize the accuracy and power of your research. *American Psychologist, 63*, 591-601. https://doi.org/10.1037/0003-066X.63.7.591

Global Deception Research Team (2006). A world of lies. *Journal of Cross-Cultural Psychology*, *37*, 60-74. https://doi.org/10.1177/0022022105282295

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. (2020). *Duped. Truth-default theory and the social science of lying and deception*. Tuscaloosa, AL: University of Alabama Press.

Luke, T. J. (2019). Lessons from Pinocchio: Cues to deception may be highly exaggerated. *Perspectives on Psychological Science, 14*, 646-671. https://doi.org/10.1177/1745691619838258

Masip, J. (2017). Deception detection: State of the art and future prospects. *Psicothema, 29*, 149-159. https://doi.org/10.7334/psicothema2017.34

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

Masip, J., & Sánchez, N. (2019). How people *really* suspect lies: A re-examination of Novotny et al.'s (2018) data. *Journal of Nonverbal Behavior, 43*, 481-492. https://doi.org/10.1007/s10919-019-00309-y

Noguchi, K., Gel, Y. R., Brunner, E., & Konietschke, F. (2012). nparLD: An R software package for the nonparametric analysis of longitudinal data in factorial experiments. *Journal of Statistical Software, 50*(12), 1-23.

Nortje, A., & Tredoux, C. (2019). How good are we at detecting deception? A review of current techniques and theories. *South African Journal of Psychology, 49*, 491-504. https://doi.org/10.1177/0081246318822953

Novotny, E., Carr, Z., Frank, M. G., Dietrich, S. B., Shaddock, T., Cardwell, M., & Decker, A. (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

Reinhard, M.-A., & Sporer, S. L. (2008). Verbal and nonverbal behavior as a basis for credibility attribution: The impact of task involvement and cognitive capacity. *Journal of Experimental Social Psychology, 44*, 477-488. https://doi.org/10.1016/j.jesp.2007.07.012

Reinhard, M.-A., & Sporer, S. L. (2010). Content versus source cue information as a basis for credibility judgments: The impact of task involvement. *Social Psychology, 41*, 93-104. https://doi.org/10.1027/1864-9335/a000014

Reinhard, M.-A., Sporer, S. L., Scharmach, M., & Marksteiner, T. (2011) Listening, not watching: Situational familiarity and the ability to detect deception. *Journal of Personality and Social Psychology, 101*, 467-484. https://doi.org/10.1037/a0023726

Rogmann, J. (2013). *orddom: Ordinal dominance statistics*. R package. Retrieved from https://cran.r-project.org/web/packages/orddom/

Sporer, S. L., & Schwandt, B. (2006). Paraverbal indicators of deception: A meta-analytic synthesis. *Applied Cognitive Psychology, 20*, 421-446. https://doi.org/10.1002/acp.1190

Sporer, S. L., & Schwandt, B. (2007). Moderators of nonverbal indicators of deception: A meta-analytic synthesis. *Psychology, Public Policy, and Law, 13*, 1-34. https://doi.org/10.1037/1076-8971.13.1.1

Stiff, J. B., Miller, G. R., Sleight, C., Mongeau, P., Garlick, R., & Rogan, R. (1989). Explanations for visual cue primacy in judgments of honesty and deceit. *Journal of Personality and Social Psychology, 56*, 555-564. https://doi.org/10.1037/0022-3514.56.4.555

Vrij, A. (2008). *Detecting lies and deceit: Pitfalls and opportunities*. Chichester, UK: Wiley.