Linguistic Concreteness of Statements of True and False Intentions

Sofia Calderon[a]

Erik Mac Giolla[a]

Timothy J. Luke[a]

Lara Warmelink[b]

Karl Ask[a]

Pär Anders Granhag[a, c]

Aldert Vrij[d]

[a]Department of Psychology, University of Gothenburg,

Gothenburg, Sweden

[b]Department of Psychology, Lancaster University,

Lancaster, United Kingdom

[c]Norwegian Police University College,

Oslo, Norway

[d]Department of Psychology, University of Portsmouth,

Portsmouth, United Kingdom

Author Note

This work was supported by the Swedish Research Council [grant number 2015-02144].

The authors would like to thank Igor Tabak for his assistance with data preparation.

Correspondence concerning this article should be addressed to Erik Mac Giolla, Department of Psychology, University of Gothenburg, SE 405 30 Gothenburg, Sweden. E-mail: erik.mac.giolla@psy.gu.se

The research questions, hypotheses, and analysis plan (including R code) were pre-registered before analyses were conducted. Study decisions and material can be found on the OSF page https://osf.io/y48cz/. Note that we did not initially plan to use the automated LCM coding as a measure in the study, but as we found an automated LCM-coding protocol during the process of writing the paper, we decided to use it to cross-validate the outcome of the study. Therefore, there are two different pre-registrations for the study.

Abstract

Our aim was to examine how people communicate their *true* and *false* intentions. Based on construal level theory (Trope & Liberman, 2010), we predicted that statements of true intentions would be more concretely phrased than statements of false intentions. True intentions refer to more likely future events than false intentions, and they should therefore be mentally represented at a lower level of mental construal. This should be mirrored in more concrete language use. Transcripts of truthful and deceptive statements about intentions from six previous experimental studies (total $N = 528$) were analyzed using two automated verbal content analysis approaches: a folk-conceptual measure of concreteness (Brysbaert et al., 2014) and linguistic category model scoring (Seih et al., 2017). Contrary to our hypotheses, veracity did not predict statements' concreteness scores, suggesting that automated verbal analysis of linguistic concreteness is not a viable deception-detection technique for intentions.

Keywords: construal level theory, true and false intentions, mental abstraction, automated text analysis, deception

General audience summary

Previous research shows that people express themselves in different ways depending on whether they are lying of telling the truth. It has for example been found that true accounts contain more sensory information and more details that can be checked or verified than false accounts. This study extends the previous literature by examining if *concreteness* can be used as a cue to deception. We relied on Construal Level Theory (CLT) which explains how people think about events that are not experienced directly, here and now. Simply put, events that are perceived to be “far away” from oneself are thought of more abstractly than events perceived as “near”. For example, an unlikely event should be perceived as further away from oneself than a likely event, and thereby thought of in more abstract terms. This should in turn be mirrored in the language people use to describe the different situations. In our study, we analyzed a large number of statements of true and false intentions—that is, lies and truths about the future—based on CLT assumptions. Specifically, since false intentions (lies) refer to unlikely events they should be thought of and expressed in more abstract terms, whereas true intentions (truths) refer to likely events and should instead be thought of and expressed in more concrete terms. In this study we re-analyzed statements collected in six previous experiments on true and false intentions. Specifically, we used two automated measures of language concreteness to assess potential differences in the language used to describe true and false intentions. The results showed no relationship between the veracity of statements and concreteness. We therefore advise against using concreteness as a cue to distinguish between statements of true and false intent.

Linguistic Concreteness of Statements of True and False Intentions

People often communicate what they intend to do in the future, such as describing the purpose of their trip to a border-control officer. Often such statements reflect what people genuinely intend to do (a *true intention*). Sometimes, however, people lie in order to mask their actually intended future behavior (a *false intention*). While such lies are mostly harmless, such as falsely claiming you intend to go to the gym this week, there are situations where the ability to detect false intentions is crucial. For example, border-control officers may prevent criminal activity by accurately judging the veracity of people's claimed reasons for entering a country (Granhag & Mac Giolla, 2014). In the current research, we apply construal level theory (Trope & Liberman, 2010) and a verbal content-analytical approach to the study of true and false intentions. Because CLT predicts that intended (vs. unintended) future behaviors are less psychologically distant, we test the hypothesis that people express true intentions in more concrete terms than their false intentions. In a mega-analysis—an approach to synthesize past studies using the raw data of the original studies (see, e.g., Scoboria et al., 2017)—we examined the concreteness of statements by automated coding of verbal data previously collected in six experiments.

Construal Level Theory

Construal level theory (CLT) concerns the mental representation of situations that are not directly experienced (Trope & Liberman, 2010). CLT suggests that people represent events and objects on a continuum from low-level to high-level, where low-level representations are concrete and complex and high-level representations are abstract and simple (Trope & Liberman, 2003). Higher mental abstraction is predicted to produce a range of cognitive effects, including broader object categorization (Henderson et al., 2006) and increased ease of processing for stimuli that have similar levels of construal (Bar-Anan et al., 2007). Construal level is also mirrored in communication (Soderberg et al., 2014).

Specifically, higher construal levels translate to more abstract language use, and lower construal levels to more concrete language use (Bhatia & Walasek, 2016).

CLT holds that psychological distance should directly influence construal level: events and objects that are psychologically distant are mentally represented at a higher, more abstract level than events that are perceived as closer to the self (Liberman & Trope, 2014). Four main types of psychological distance have been proposed to affect mental abstraction: temporal, spatial, social, and hypothetical. Liberman et al. (2002) found that participants who imagined engaging in an event grouped objects related to the events into fewer (more inclusive) categories when it was to take place in a year's time (distant future) compared to the upcoming weekend (near future). Others have found similar results for the other distances: higher levels of mental construal are associated with events that take place in geographically more distant locations (Fujita et al., 2006), that happen to someone less similar to oneself (Liviatan et al., 2008), and that are perceived to be less likely to occur (Wakslak et al., 2006).

Extending Construal Level Theory to Deception Contexts

These insights from construal level theory can be used to inform predictions about true and false intentions. A true intention can be defined as a claimed future action genuinely intended to be performed, whereas a false intention is instead a claimed future action not intended to be performed (Granhag, 2010). Since a true intention refers to a genuinely intended action, it should be perceived by the intender as a high-likelihood future event.[1] According to CLT, this should result in a concrete low-level mental construal. A false intention, on the other hand, refers to a claimed but not intended future action and should therefore constitute a low-likelihood future event, and should result in an abstract high-level mental construal. Insofar as construal level is mirrored in communication (Soderberg et al., 2014), it follows that true intentions should be described using more concrete language compared to false intentions.

Although our hypothesis was inspired by CLT, the idea that statements of true intent should be more concrete than statements of false intent dovetails well with other theories and findings within deception research. Indeed, Kleinberg et al. (2019) highlight three frameworks suggesting that truths should be more concrete than lies: Reality Monitoring (RM; Masip et al., 2005), Criteria Based Content Analysis (CBCA; Undeutsch, 1989), and the verifiability approach (Nahari et al., 2014). Taken together, these frameworks suggest that truthful accounts should contain more sensory information, more contextual information, and more details that can be checked or verified. It seems reasonable that accounts containing more of these types of details should also be more concrete (but see Mac Giolla et al., 2017; Mac Giolla et al., 2019, for a critical discussion on extending these theories to the context of true and false intentions).

To our knowledge, the only other study to have examined differences in concreteness between true and false statements of intent is Kleinberg et al. (2019). They used 8 measures of concreteness on true and false written statements from six datasets, three of which examined true and false intentions. Truth tellers were rated as more concrete in two out of the three intentions datasets, using some measures, but not others. When averaging across the three datasets they found a small effect of concreteness. A limitation of Kleinberg et al. is that they only analyzed written statements. We extend this work by including both spoken and written statements. Importantly, spoken communication has been shown to be more concrete than written communication (DeVito, 1967).

The Current Study

Our aim was to examine whether statements of true and false intent differ in terms of concreteness as estimated by automated measures of concreteness (for an overiew of automated verbal approaches, see Hauch et al., 2014). We analyze datasets from six previous empirical studies on true and false intentions. We used two measures of language abstraction:

(a) the folk-conceptual dictionary (Brysbaert et al., 2014), and (b), an automated version of the Linguistic Category Model (LCM; Semin & Fiedler, 1991) developed by Seih et al. (2017). Both dictionaries have successfully uncovered language differences in line with CLT (Bhatia & Walasek, 2016; Seih et al., 2017).

Hypotheses

We had three hypotheses. First, we predicted that truth tellers' statements would be more concrete than liars' statements (H1). Second, we considered the effect statement length has on the effect of veracity on concreteness score. We reasoned that truth tellers who give longer statements will add concrete information to their statements (vs. Truth tellers with shorter statements), while liars who give longer statements may add mostly abstract information. Therefore, we predicted that a greater difference in concreteness between truth tellers and liars when statements are longer (H2). Third, since some questions in the material (e.g., "where are you flying today?") constrain the potential variability in concreteness scores, they may reduce the room for differences between liars and truth tellers to emerge, diluting the difference predicted in H1. Therefore, we predicted greater differences between liars and truth tellers in concreteness scores when statements included more variability in concreteness (H3). In addition to the three hypotheses, we explored which interview questions produced the greatest differences between truth tellers and liars in concreteness scores.

Method

Data Characteristics

We obtained transcripts of answers to interview questions from six previous empirical studies on true and false intentions (*N* = 528 participants; see Table 1 for more information about the separate studies). Our inclusion criteria were: Experimental studies, conducted in English, including statements of both true and false intentions. Two additional studies (Fenn et al., 2015; Mann et al., 2012) also met our inclusion criteria, but we were unable to obtain the raw data from the authors.

Each study included between 4 and 25 interview questions, for a total of 90 unique questions in the aggregated dataset. These touched upon different themes, such as open-ended questions asking participants to describe their whole intention, specific questions regarding time and place for their claimed future plans, and questions about whether and how participants had planned for their allegedly intended behavior (see Table S1 at https://osf.io/y48cz/ for the complete set of questions; see Appendix A for manipulation instructions used in the studies included in the mega-analysis). Five studies had a between-subjects design, and one had a within-subjects design.

Table 1

Characteristics of Studies Included in the Mega-analysis

Study	Setting	*n*	Mean age (SD)	Percentage male	Description
Warmelink, Vrij, Mann, & Granhag (2013, Exp. 2)	Field	84	58.0 (12.6)	57.1	Passengers on a ferry either told the truth or lied about their upcoming trip
Warmelink, Vrij, Mann, Jundi, and Granhag (2012)	Lab	86	27.6 (12.3)	23.3	Participants told the truth or lied about an upcoming trip
Warmelink (2017)	Lab	45	21.0 (5.58)	20	Participants told the truth or lied about a trip/last day of term
Kleinberg, Nahari, Arntz, & Verschure (2017)	Online	222	33.6 (9.07)	[N/A]	Participants told the truth or lied about a trip
Vrij, Granhag, Mann, and Leal (2011)	Field	60	45.2 (12.8)	66.7	Passengers at an airport told the truth or lied about their upcoming trip

Vrij, Leal, Mann, and Granhag (2011)	Lab	31	36.1 (8.21)	90.6	Participants honestly or deceptively described their mission

Measures

The folk-conceptual approach (Brysbaert et al., 2014) uses a dictionary for automated coding of texts for concreteness. It consists of 40,000 common English words, each of which has been rated by on average 30 participants (over 4,000 participants overall) on a scale from 1 (abstract) to 5 (concrete). To illustrate the range of concreteness scores, examples at the extremes include “essentialness” (M = 1.04) and “sea turtle” (M = 5.00). Words with intermediate concreteness include “notify” (M = 3.04) and “gossip” (M = 3.00). For the folk-concreteness measure, words in the dictionary were matched with each word in the statements, and mean scores for each statement were calculated.

The LCM (Semin & Fiedler, 1991) represents a more theory-driven approach to examining language abstraction. The method relies on the assumption that certain classes of verbs and adjectives vary in their level of abstraction. Categories involving emotional states or enduring dispositional traits are considered more abstract, whereas specific and easily imaginable actions are considered more concrete. In order of ascending concreteness, the LCM coding procedure counts the occurrence of nouns (e.g., “harasser”), adjectives (e.g., “aggressive”), state verbs (e.g., “hate”), interpretative action verbs (e.g., “hurt”), and descriptive action verbs (e.g., “hit”; Coenen et al., 2006). For the LCM, an automated part-of-speech tagger called TreeTagger (Schmid, 1995) was used to identify the five types of language categories. In line with the procedures of Seih et al. (2017), each category received a weight from 1 (the most concrete category ‘descriptive action verbs’) to 5 (the most abstract category ‘nouns’), based on which a mean score was calculated for each statement.

Data Preparation and Analysis

To account for the random effects of participants, studies, and questions, we chose a mixed-effects (multi-level) approach to analyzing the data. The transcribed answers to the interview questions were entered in a document along with indicators of veracity condition, study name, participant ID, question ID, and a question label.[2] The primary dependent variable of concreteness was computed when all raw data had been entered. There was a total of 6,104 observations (i.e., statements): 3,005 true statements and 3,106 false statements. The folk-concreteness scores ranged from 1.43 to 4.96 and LCM scores ranged from 1.00 to 5.00. Statement length was also calculated and varied from 1 to 1,416 words (M = 32.6, Mdn = 18.0).

To test our hypotheses, we performed a series of linear mixed-effects analyses using R (R Core Team, 2013) and the *lme4* package (Bates et al., 2014). We also used the *lmerTest* package (Kuznetsova et al., 2017) to calculate Satterthwaite-approximated degrees of freedom (Gaylor & Hopper, 1969) and corresponding p-values to assess the significance of the coefficients. We modeled random intercepts for each study, question, and subject and random slopes for each subject for the effect of veracity. To estimate the support for the null hypothesis relative to the alternative, we also calculated a Bayes Factor (BF) from the Bayesian Information Criterion (BIC) using the method outlined by Wagenmakers (2007). We report BIC Bayes factors in favor of the null (i.e., initial models without the predictors of interest) for all the models.

Results

Table 2

Descriptive Statistics for Concreteness Scores over Veracity Conditions

	Folk concreteness[a]	LCM[b]

	True intention	False intention	True intention	False intention
Mean	2.55	2.52	3.75	3.78
Median	2.53	2.49	3.67	3.70
Standard deviation	0.35	0.34	0.55	0.54

[a] Higher values indicate more concrete language

[b] Lower values indicate more concrete language

Folk-Concreteness Score

To test if truth tellers' statements were more concrete than liars' statements (H1), we specified a model predicting the folk-concreteness score of statements from veracity entered as a fixed effect. Study, subject (nested within study), and question (nested within subject) were entered as random effects. Failing to provide support for H1, veracity did not significantly predict participants' concreteness score, $b = -0.04$, 95% CI [−0.07, −0.01], $t(3.21) = -2.87$, $p = .059$.[3, 4, 5] There was a statistically non-significant mean difference in concreteness score of 0.04 in the predicted direction. The model's estimates of the variance components are also informative. The random intercepts for subjects varied from 2.40 to 2.72 (range .32). The random slopes for subjects varied from −0.02 to −0.06 (range .04). Thus, the individual variation in linguistic concreteness was substantially greater than the effect of veracity.

To calculate a Bayes factor, we specified the null model as our initial model without the veracity factor. The analysis revealed strong support in favor of the null hypothesis, $BF_{01} = 96$. See Table 2 for descriptive statistics and Table 3 for examples of texts with high and low levels of abstraction[6]. Figure 1 depicts distributions and box plots for the concreteness scores in the truthful and deceptive conditions.

Table 3

Example texts of with high and low levels of abstraction

Concreteness[a]	**Condition**	**Example text**
1.66	Truthful	"There hasn't been anything that's been very hard."
3.79	Truthful	"Baggage missing, plane crash, death, car accident, birds shitting on my head, poison in my food."
1.61	Deceptive	"It's like 11-ish because we've probably had drinks, so..."
3.99	Deceptive	"Er, I'm flying back on, err, Sunday. Sunday night I'm flying back, so..."

[a]Scale ranges from 1 (abstract) to 5 (concrete). Observed range 1.43 to 4.96.

To test whether statement length moderates the difference in concreteness between liars and truth tellers' statements (H2), we formulated a model predicting the concreteness score of statements from veracity, statement length, and an interaction term for veracity by statement length. Study, subject (nested within study), and question (nested within subject) were again entered as random effects. Veracity did not predict concreteness score, $b = -0.04$, 95% CI [−0.07, −0.01], $t(3.20) = -2.77$, $p = .064$, $BF_{01} = 43{,}231{,}950$, and neither did the interaction between statement length and veracity, $b = -2.17 \times 10^{-4}$, 95% CI [-1.06×10^{-4}, 5.40×10^{-4}], $t(1004) = -1.32$, $p = .188$, $BF_{01} = 80{,}495$. Length, however, did significantly predict concreteness score, $b = -2.49 \times 10^{-4}$, 95% CI [3.66×10^{-5}, 4.61×10^{-4}], $t(2522) = -2.30$, $p = .022$, $BF_{01} = 8{,}151{,}846$. This indicates that longer texts were less concrete, which goes against our predictions. However, it should be noted that this effect was very small in absolute terms: A 100-word increase in statement length was associated with a 0.02-unit decrease in concreteness. Furthermore, the Bayes factors indicate extreme support for the null model. Specifically, the data are more than 8 million times more likely under a null model (specified as the initial model without statement length and the interaction term) than under the alternative model (veracity as sole predictor of concreteness score).

We also tested whether questions that elicit more variability in concreteness scores elicit greater differences between liars and truth tellers (H3). We predicted the standardized mean difference between veracity conditions on concreteness score for each question (Cohen's *d*) from the standard deviation of concreteness scores for responses to each question. Variability in concreteness scores did not significantly predict veracity effects on concreteness scores, $b = 0.25$, 95% CI [−0.23, 0.73], $t(3.72) = 1.01$, $p = .372$, $BF_{01} = 9.62$. As an additional exploratory examination of this hypothesis, we fit a model predicting the unstandardized mean difference with the standard deviation of concreteness scores. In this model, variability in concreteness scores again did not significantly predict veracity effects on concreteness scores, $b = 0.26$, 95% CI [−0.11, 0.63], $t(4.49) = 1.86$, $p = .13$, , $BF_{01} = 5.93$.

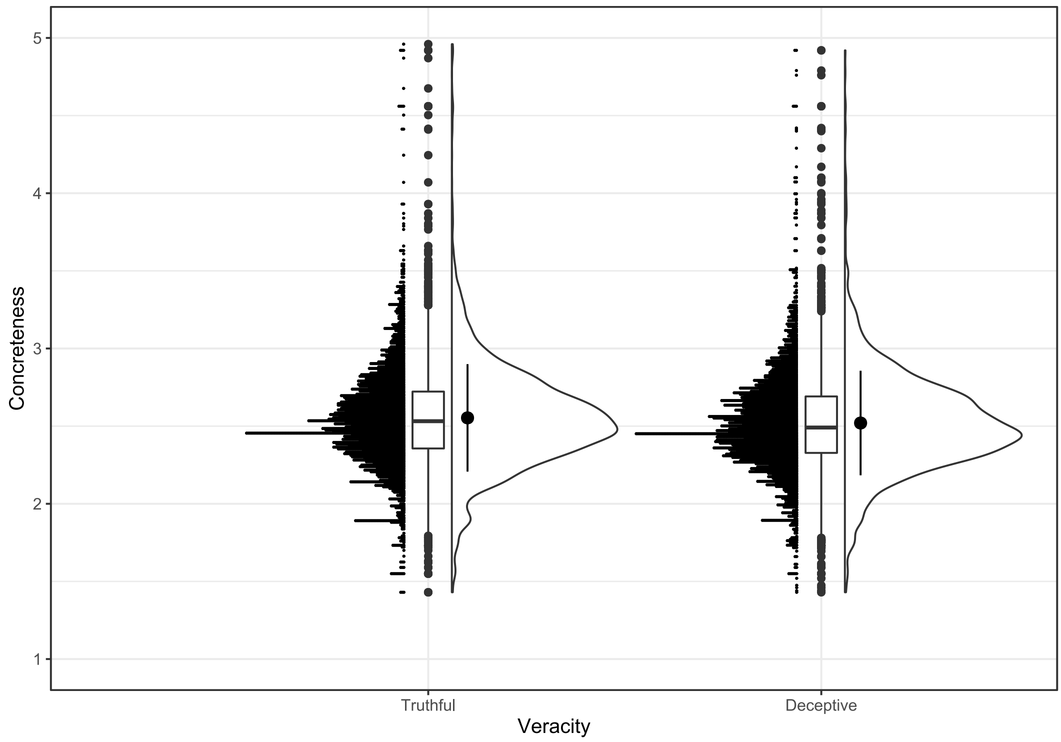

Figure 1. Histograms, box plots, 95% CI of the means, and density distributions for the Folk-concreteness scores of the 3,005 true statements and 3,106 deceptive statements. Note the near identical distributions for truth tellers and liars.

To explore which specific questions produce the greatest differences between truth tellers and liars, we created a forest plot of the effect sizes (Hedges's *g*) for each interview question (see Figure 2), using the *metafor* package for R (Viechtbauer, 2010). A couple of observations should be noted. First, the effects are approximately evenly distributed both below and above zero. Second, only seven out of the ninety interview questions resulted in statistically significant effects in the predicted direction, and two effects were statistically significant in the opposite direction. The seven interview questions that produced effects in the predicted direction were:

- What is the main purpose of your flight to X? (Kleinberg et al, 2017; standard condition)
- Please describe in as much detail as possible all the steps you had to take in preparation of this trip. (Kleinberg et al, 2017; specific condition)
- Please could you describe in as much detail as possible how you will get from here to that location? (Vrij, Leal, Mann, & Granhag, 2011)
- Are you intending to do any preparation of planning in the future to execute your intention? (Warmelink et al, 2017)
- What is the main purpose of your trip? (Warmelink, Vrij, Mann, Juni, & Granhag, 2013)
- What is the most important place you'll go on your trip? (Warmelink, Vrij, Mann, Juni, & Granhag, 2013)
- How long will the journey take? (Warmelink, Vrij, Mann, Juni, & Granhag, 2013)

The two interview questions that produced effects in the opposite direction were both from the "specific" condition in the experiment by Kleinberg et al (2017):

- Please describe in which order you did the planning for your trip to X? What was first, what second, what was last?
- What is the most unpleasant event you expect to happen during your trip?

To avoid capitalizing on chance findings, we do not want to speculate about reasons why these specific questions rendered statistically significant effects. The present data cannot determine whether these results reflect a systematic pattern or mere random variation.

Linguistic Category Model Scores

In line with the results for the folk-conceptual measure, the results from the LCM-analyses did not support our hypotheses. Veracity did not significantly predict participants' concreteness score, $b = 0.02$, 95% CI [−0.02, 0.06], $t(11.1) = 0.77$, $p = .456$, $BF_{01} = 1,068$, lending strong support for the null hypothesis as opposed to H1. Adding statement length and the Length × Veracity interaction term to the model did not provide any support for H2: Again, veracity did not predict concreteness score, $b = 0.02$, 95% CI [−0.02, 0.06], $t(13.6) = 0.82$, $p = .426$, $BF_{01} = 70,103,627$, neither did length, $b = -2.26 \times 10^{-4}$, 95% CI [-7.64×10^{-5}, 5.29×10^{-4}], $t(2779) = 1.47$, $p = .143$, $BF_{01} = 22,434,475$, or the interaction of statement length and veracity, $b = -4.66 \times 10^{-4}$, 95% CI [-9.32×10^{-4}, -1.96×10^{-7}], $t(886) = 1.96$, $p = .050$, $BF_{01} = 21,089$. Failing to support H3, variability in concreteness scores did not significantly predict veracity effects on concreteness scores, although the Bayes factor suggest that the evidence for the null is weak, $b = 0.52$, $t(4.76) = 1.81$, $p = .133$, $BF_{01} = 2.92$.

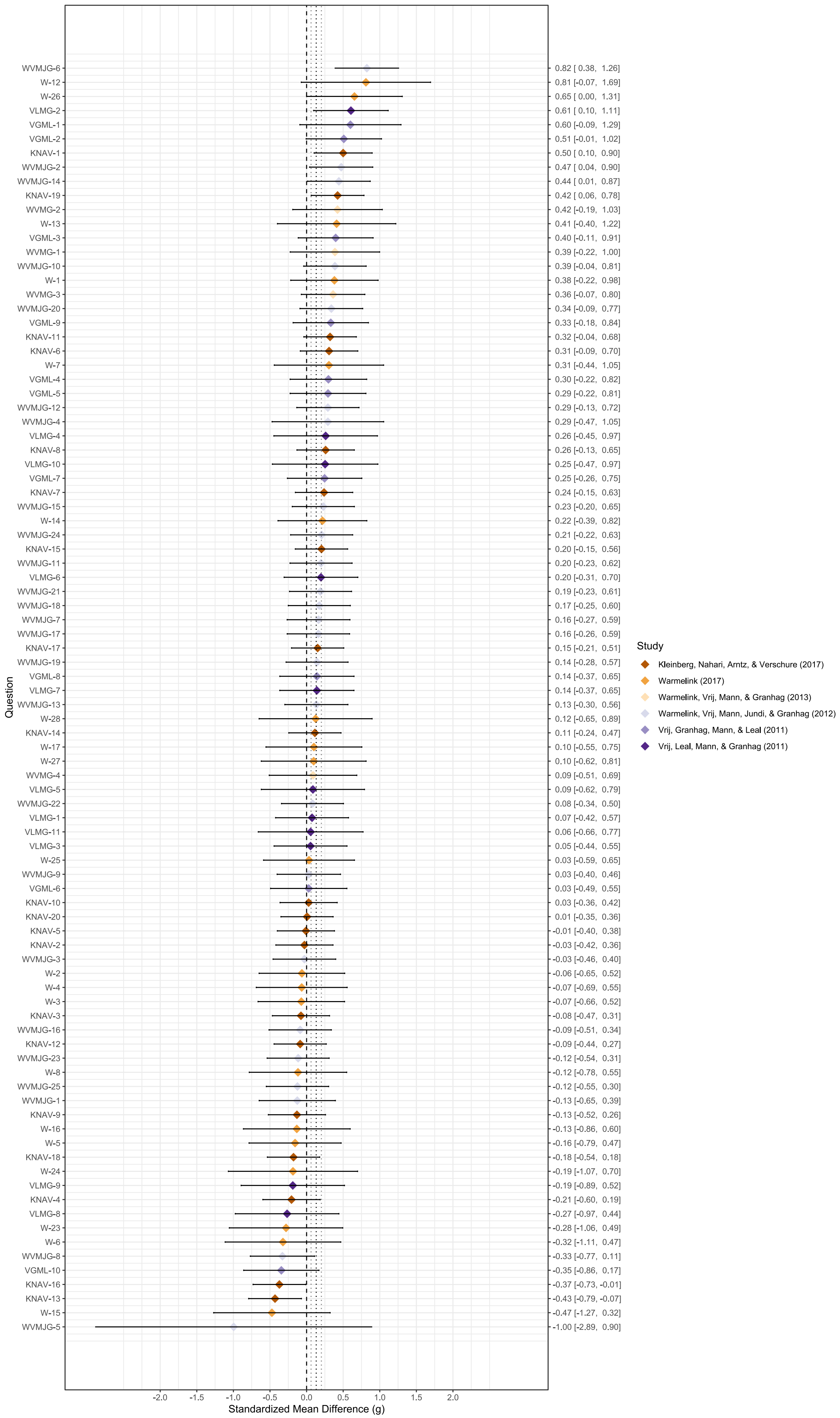
Question
Standardized Mean Difference (g)
Study
Kleinberg, Nahari, Arntz, & Verschure (2017)
Warmelink (2017)
Warmelink, Vrij, Mann, & Granhag (2013)
Warmelink, Vrij, Mann, Jundi, & Granhag (2012)
Vrij, Granhag, Mann, & Leal (2011)
Vrij, Leal, Mann, & Granhag (2011)

Figure 2. Forest plot over effect sizes (Hedges' *g*) for each interview question based on the folk-concreteness score, a legend for each question is in Table S1 at https://osf.io/y48cz/.

In addition to our preregistered analyses, we also explored the correlation between the folk-concreteness scores and the LCM scores. On the assumption that both coding systems measure concreteness, we expected a substantial correlation between the two. Unexpectedly, however, there was a near-zero correlation between the folk-concreteness scores and the LCM scores ($r = .025$, $p = .050$). This suggests that one or both of the coding systems may lack conceptual validity and/or reliability[7].

Discussion

We tested the prediction that truth tellers would use more concrete language than liars when communicating their true and false intentions. The study provided no empirical support for any of our hypotheses: statements of true intentions were not more concretely phrased than statements of false intentions (H1); statement length did not moderate an effect of veracity on differences in concreteness (H2); and questions that elicit more variability in concreteness scores did not elicit greater differences between liars and truth tellers (H3). In addition, specific types of questions did not systematically influence concreteness.

There are several reasons why we believe it is unlikely that our null results are false negatives. First, the large number of observations (over 6,000 statements) as well as Bayes factors in strong support of the null hypotheses gives us confidence in the strength of the statistical analyses. Second, the null findings are consistent across multiple studies, questions, and designs. Third, our results for one of the concreteness measures—the folk-concreteness measure—are corroborated by Kleinberg et al. (2019). However, while we found no differences in true and false statements when using the LCM measure, Kleinberg et al. found a difference in two of their three datasets. This discrepancy requires further research. This difference may be due to differences in the exact datasets included in the analyses, such as Kleinberg et al.'s focus on written statements compared to the greater variety of statements in our dataset. There may also be currently unknown moderators of a veracity effect. Considering the relatively small effect size found in Kleinberg et al. (meta-analytic Cohen's d = 0.28), the possibility of either a false positive in Kleinberg et al. or a false negative on the LCM measure in the current study is also substantial.

There are several potential explanations for our own null findings. First, perhaps true and false intentions do differ in construal level, but the automated measures of linguistic abstractness failed to capture it. There are inherent problems with simple word-count methods (see limitations). Calderon et al. (2018) found that drawings of false intentions were perceived as more abstract than drawings of true intentions, which suggests that other measures may capture differences in abstraction in line with CLT.

Second, aspects other than psychological distance may have influenced the level of language abstraction. People purposely tailor their messages to suit a desired outcome. For example, the goal of being liked by a recipient influences level of linguistic concreteness (Rubini & Sigall, 2002). Liars frequently strive to be liked, while expecting to be disbelieved. Such goals and expectations may have acted as suppressors of an effect of psychological distance on mental abstraction in the current study.

Third, true and false intentions may not readily translate to high- and low-likelihood future events. If there is no difference in likelihood, we should not expect differences in mental abstraction. Liars often prepare a cover story—a verbal lie-script—to use when being questioned about their plans (Granhag et al., 2015). This mental preparation may decrease the psychological distance to the stated intention. If so, the assumption that true and false intentions differ in psychological distance may not be valid, which would render linguistic abstraction non-diagnostic in the specific context of deception.

A final potential explanation is that the fundamental assumption in CLT, that subjective likelihood influences construal level, does not hold true. In two replication attempts Wakslak et al. (2006), Calderon et al. (2020) failed to replicate the basic finding that people construe future events more abstractly under low-likelihood conditions compared to high-likelihood conditions. Since Wakslak et al. (2006) constitutes the primary empirical work for the predictions tested in the current study, the unreliability of the original finding may explain the current null-findings. Relatedly, it is also possible, despite this being one of CLTs fundamental assumptions, that construal levels do not influence levels of abstraction. Maier et al. (2022) used novel bias correction methods to reanalyse the data underlying the largest meta-analysis on construal level theory (Soderberg et al., 2014). Using a Robust Bayesian Meta-Analysis (RoBMA; Maier et al., 2020), they found evidence against an effect of construal level on abstraction.

Limitations

One limitation with the current study is the control that the interviewer exerted over the interviews in which the statements were produced. Perhaps in a completely free narrative, true statements of intentions would be more concrete than false statements of intentions. The fact that different questions elicit such varying levels of concreteness suggests that the manner in which responses are elicited plays a significant role. However, many of the questions that were asked initially during the interviews were open ended (e.g., “please describe your whole intention”). If the null findings were dependent on this aspect of the interview styles, systematic differences between open- and closed-ended questions should be expected. No such pattern was observed.

Another limitation concerns the dictionaries we used to measure linguistic concreteness. Because these include a limited word set, there is some lack of precision in our analyses. This may have limited our ability to detect subtle differences in concreteness between truthful and deceptive statements. It is also possible that automated computer analysis of concreteness is not yet sufficiently developed for this purpose. One way around the limitation of simple word-count methods is to use concreteness ratings of multiword expressions, which have recently been developed (Muraki et al., 2022). Another option is to use human coders. In the so-called best–worst scoring approach (Hollis, 2018), pairs of texts are presented to judges who make relative judgments on some dimension of interest, such as concreteness. Future studies could explore the usefulness of these methods.

Conclusions

The goal of this study was to test whether true statements of intentions are more concretely expressed than false statements of intentions. Our mega-analysis of experimental data did not provide any empirical support for this prediction. Instead, true and false intentions seem to be expressed at similar levels of concreteness. Employing the automated measures of linguistic abstraction tested here is not a viable technique to detect deception about intentions.

References

Bar-Anan, Y., Liberman, N., Trope, Y., & Algom, D. (2007). Automatic processing of psychological distance: evidence from a Stroop task. *Journal of Experimental Psychology: General, 136*(4), 610. Doi:10.1037/0096-3445.136.4.610

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). Lme4: Linear mixed-effects models using Eigen and S4. *R package version, 1*(7), 1-23.

Bhatia, S., & Walasek, L. (2016). Event construal and temporal distance in natural language. *Cognition, 152*, 1-8. Doi:10.1016/j.cognition.2016.03.011

Brysbaert, M., Warriner, A. B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods, 46*(3), 904-911. Doi:10.3758/s13428-013-0403-5

Calderon, S., Ask, K., Mac Giolla, E., & Granhag, P. A. (2019). *The Mental Representation of* True and False Intentions: A Comparison of Schema-Consistent and Schema-Inconsistent Tasks. *Unpulished manuscript.*

Calderon, S., Mac Giolla, E., Ask, K., & Granhag, P. A. (2018). Drawing what lies ahead: False intentions are more abstractly depicted than true intentions. *Applied Cognitive Psychology, 32*(4), 518-522. doi:10.1002/acp.3422

Calderon, S., Mac Giolla, E., Ask, K., & Granhag, P. A. (2020). Subjective likelihood and the construal level of future events: A replication study of Wakslak, Trope, Liberman, and Alony (2006). *Journal of Personality and Social Psychology*. Advance online publication. doi:10.1037/pspa0000214

Calderon, S., Mac Giolla, E., Granhag, P. A., & Ask, K. (2017). Do true and false intentions differ in level of abstraction? A test of Construal level theory in deception contexts. *Frontiers in Psychology, 8*(2037). doi:10.3389/fpsyg.2017.02037

Carlson, M., & Miller, N. (1987). Explanation of the relation between negative mood and helping. *Psychological Bulletin*, *102*(1), 91-108. doi:10.1037/0033-2909.102.1.91

Coenen, L. H., Hedebouw, L., & Semin, G. R. (2006). *The linguistic category model (LCM) manual*. Unpublished manuscript.

DePaulo, B. M., Lindsay, J. J., Malone, B. E., Muhlenbruck, L., Charlton, K., & Cooper, H. (2003). Cues to deception. *Psychological Bulletin, 129*(1), 74-118. doi:10.1037/0033-2909.129.1.74

DeVito, J. A. (1967). Levels of abstraction in spoken and written language. *Journal of Communication*, *17*(4), 354-361. https://doi.org/10.1111/j.1460-2466.1967.tb01193.x

Fenn, E., McGuire, M., Langben, S., & Blandón-Gitlin, I. (2015). A reverse order interview does not aid deception detection regarding intentions. *Frontiers in Psychology, 6*, 1298. doi:10.3389/fpsyg.2015.01298

Fujita, K., Henderson, M. D., Eng, J., Trope, Y., & Liberman, N. (2006). Spatial distance and mental construal of social events. *Psychological Science, 17*(4), 278-282. doi:10.1111/j.1467-9280.2006.01698.x

Gaylor, D., & Hopper, F. (1969). Estimating the degrees of freedom for linear combinations of mean squares by Satterthwaite's formula. *Technometrics, 11*(4), 691-706. doi:10.2307/1266893

Gibson, J. W., Gruner, C. R., Kibler, R. J., & Kelly, F. J. (1966). A quantitative examination of differences and similarities in written and spoken messages. *Communications Monographs*, *33*(4), 444-451. https://doi.org/10.1080/03637756609375510

Gilead, M., Liberman, N., & Maril, A. (2012). Construing counterfactual worlds: The role of abstraction. *European Journal of Social Psychology, 42*(3), 391-397. doi:10.1002/ejsp.1862

Granhag, P. A. (2010). On the psycho-legal study of true and false intentions: Dangerous waters and some stepping stones. *The Open Criminology Journal, 3*, 37-43. doi:10.2174/1874917801003010037

Granhag, P. A., Hartwig, M., Mac Giolla, E., & Clemens, F. (2015). Suspects' verbal counter-interrogation strategies: Towards an integrative model. In P. A. Granhag, A. Vrij, & B. Verschuere (Eds.), *Deception detection: Current challenges and new approaches* (pp. 293-313). Chichester: John Wiley.

Granhag, P. A., & Mac Giolla, E. (2014). Preventing future crimes: Identifying markers of true and false intent. *European Psychologist, 19*(3), 195-206. doi:10.1027/1016-9040/a000202

Hauch, V., Blandón-Gitlin, I., Masip, J., & Sporer, S. L. (2014). Are computers effective lie detectors? A meta-analysis of linguistic cues to deception. *Personality and Social Psychology Review*. doi:10.1177/1088868314556539

Henderson, M. D., Fujita, K., Trope, Y., & Liberman, N. (2006). Transcending the "here": the effect of spatial distance on social judgment. *Journal of Personality and Social Psychology, 91*(5), 845-856. doi:10.1037/0022-3514.91.5.845

Hollis, G. (2018). Scoring best-worst data in unbalanced many-item designs, with applications to crowdsourcing semantic judgments. Behavior research methods, 50(2), 711-729.

Iliev, R., Dehghani, M., & Sagi, E. (2015). Automated text analysis in psychology: Methods, applications, and future developments. *Language and Cognition, 7*(2), 265-290. doi:10.1017/langcog.2014.30

Johnson, M. K., & Raye, C. L. (1981). Reality monitoring. *Psychological Review, 88*(1), 67-85. doi:10.1037/0033-295x.88.1.67

Jupe, L. M., Leal, S., Vrij, A., & Nahari, G. (2017). Applying the Verifiability Approach in an international airport setting. *Psychology, Crime & Law, 23*(8), 812-825. doi:10.1080/1068316X.2017.1327584

Kleinberg, B., Nahari, G., Arntz, A., & Verschuere, B. (2017). An investigation on the detectability of deceptive intent about flying through verbal deception detection. *Collabra: Psychology, 3*(1). doi:10.1525/collabra.80

Kleinberg, B., van der Vegt, I., Arntz, A., & Verschuere, B. (2019, March 1). Detecting deceptive communication through linguistic concreteness. Retrieved from https://doi.org/10.31234/osf.io/p3qjh

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: tests in linear mixed effects models. *Journal of Statistical Software, 82*(13). doi:10.18637/jss.v082.i13

Liberman, N., Sagristano, M. D., & Trope, Y. (2002). The effect of temporal distance on level of mental construal. *Journal of Experimental Social Psychology, 38*(6), 523-534.

Liberman, N., & Trope, Y. (2014). Traversing psychological distance. *Trends in Cognitive Sciences, 18*(7), 364-369. doi: 10.1016/j.tics.2014.03.001

Liviatan, I., Trope, Y., & Liberman, N. (2008). Interpersonal similarity as a social distance dimension: Implications for perception of others' actions. *Journal of Experimental Social Psychology, 44*(5), 1256-1269. doi:http://dx.doi.org/10.1016/j.jesp.2008.04.007

Maass, A. (1999). Linguistic intergroup bias: Stereotype perpetuation through language. *Advances in Experimental Social Psychology* (Vol. 31, pp. 79-121): Elsevier. doi:10.1016/S0065-2601(08)60272-5

Mac Giolla, E., Ask, K., Granhag, P. A., & Karlsson, A. (2019). Can Reality Monitoring criteria distinguish between true and false intentions? *Journal of Applied Research in Memory and Cognition*, *8*(1), 92-97.

Mac Giolla, E., & Granhag, P. A. (2014). Detecting false intent amongst small cells of suspects: Single versus repeated interviews. *Journal of Investigative Psychology and Offender Profiling*, n/a-n/a. doi:10.1002/jip.1419

Mac Giolla, E., & Granhag, P. A. (2015). Detecting false intent amongst small cells of suspects: Single versus repeated interviews. *Journal of Investigative Psychology and Offender Profiling*, 12, 142-157. doi:10.10002/jip.1419

Mac Giolla, E., Granhag, P. A., & Ask, K. (2017). A goal-activation framework of true and false intentions. *Applied Cognitive Psychology, 31*(6), 678-684. doi:10.1002/acp.3366

Maier, M., Bartoš, F., Oh, M., Wagenmakers, E., Shanks, D., & Harris, A. J. L. (2022, March 3). *Publication Bias in Research on Construal Level Theory*. https://doi.org/10.31234/osf.io/r8nyu

Maier, M., Bartoš, F., & Wagenmakers, E. (2020, June 23). *Robust Bayesian Meta-Analysis: Addressing Publication Bias with Model-Averaging*. https://doi.org/10.31234/osf.io/u4cns

Mann, S., Vrij, A., Nasholm, E., Warmelink, L., Leal, S., & Forrester, D. (2012). The direction of deception: Neuro-Linguistic Programming as a lie detection tool. *Journal of Police and Criminal Psychology, 27*(2), 160-166. doi:10.1007/s11896-011-9097-8

Muraki, E. J., Abdalla, S., Brysbaert, M., & Pexman, P. M. (2022, March 2). *Concreteness ratings for 62 thousand English multiword expressions*. https://doi.org/10.31234/osf.io/m397u

Nezlek, J. B. (2008). An introduction to multilevel modeling for social and personality psychology. Social and Personality Psychology Compass, 2(2), 842-860.

Puddle-Ducks (2019). *There might be problems with the automated scoring of linguistic concreteness.* Rabbit Tracks. Retrieved from https://www.rabbitsnore.com/2019/02/there-might-be-problems-with-automated.html

R Core Team. (2013). R: A language and environment for statistical computing.

Rubini, M., & Sigall, H. (2002). Taking the edge off of disagreement: linguistic abstractness and self-presentation to a heterogeneous audience. *European Journal of Social Psychology, 32*(3), 343-351. doi:10.1002/ejsp.94

Schmid, H. (1995). Treetagger| a language independent part-of-speech tagger. *Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart*, *43*, 28.

Scoboria, A., Wade, K. A., Lindsay, D. S., Azad, T., Strange, D., Ost, J., & Hyman, I. E. (2017). A mega-analysis of memory reports from eight peer-reviewed false memory implantation studies. *Memory*, 25(2), 146-163.

Seih, Y.-T., Beier, S., & Pennebaker, J. W. (2017). Development and examination of the linguistic category model in a computerized text analysis method. *Journal of Language and Social Psychology, 36*(3), 343-355. doi:10.1177/0261927X16657855

Semin, G. R., & Fiedler, K. (1991). The Linguistic Category Model, its Bases, Applications and Range. *European Review of Social Psychology, 2*(1), 1-30. doi:10.1080/14792779143000006

Snefjella, B., & Kuperman, V. (2015). Concreteness and psychological distance in natural language use. *Psychological Science, 26*(9), 1449-1460. doi:10.1177/0956797615591771

Soderberg, C. K., Callahan, S. P., Kochersberger, A. O., Amit, E., & Ledgerwood, A. (2014). The effects of psychological distance on abstraction: Two meta-analyses. *Psychological Bulletin*, *3*(525-528). doi:10.1037/bul0000005

Sooniste, T., Granhag, P. A., Knieps, M., & Vrij, A. (2013). True and false intentions: Asking about the past to detect lies about the future. *Psychology, Crime & Law, 19*(8), 673-685. doi:10.1080/1068316x.2013.793333

Sternberg, K. J., Baradaran, L. P., Abbott, C. B., Lamb, M. E., & Guterman, E. (2006). Type of violence, age, and gender differences in the effects of family violence on children's behavior problems: A mega-analysis. *Developmental Review*, *26*, 89-112.

Stewart, L. A., & Cochrane Working Group On Meta-Analysis Using Individual Patient Data. (1995). Practical methodology of meta-analyses (overviews) using updated individual patient data. *Statistics in medicine*, *14*, 2057-2079.

Sung, Y. J., Schwander, K., Arnett, D. K., Kardia, S. L., Rankinen, T., Bouchard, C., ... & Rao, D. C. (2014). An empirical comparison of meta-analysis and mega-analysis of individual participant data for identifying gene-environment interactions. *Genetic epidemiology*, *38*(4), 369-378.

Trope, Y., & Liberman, N. (2003). Temporal construal. *Psychological Review, 110*(3), 403. doi:10.1037/0033-295X.110.3.403

Trope, Y., & Liberman, N. (2010). Construal-level theory of psychological distance. *Psychological Review, 117*(2), 440-463. doi:10.1037/a0018963

Wagenmakers, E.-J. (2007). A practical solution to the pervasive problems of p-values. *Psychonomic Bulletin & Review, 14*(5), 779-804. doi:10.3758/BF03194105

Wakslak, C. J., Trope, Y., Liberman, N., & Alony, R. (2006). Seeing the forest when entry is unlikely: Probability and the mental representation of events. *Journal of Experimental Psychology: General, 135*(4), 641-653. doi:10.1037/0096-3445.135.4.641

Warmelink, L., Vrij, A., Mann, S., & Granhag, P. A. (2013). Spatial and temporal details in intentions: A cue to detecting deception. *Applied Cognitive Psychology, 27*(1), 101-106. doi:10.1002/acp.2878

Warmelink, L., Vrij, A., Mann, S., Jundi, S., & Granhag, P. A. (2012). The effect of question expectedness and experience on lying about intentions. *Acta Psychologica, 141*(2), 178-183. doi:10.1016/j.actpsy.2012.07.011

Viechtbauer, W. (2010). Metafor: meta-analysis package for R. *R package version, 2010*, 1-0.

Vrij, A., Granhag, P. A., Mann, S., & Leal, S. (2011). Outsmarting the Liars: Toward a Cognitive Lie Detection Approach. *Current Directions in Psychological Science, 20*(1), 28-32. doi:10.1177/0963721410391245

Vrij, A., Leal, S., Mann, S. A., & Granhag, P. A. (2011). A comparison between lying about intentions and past activities: Verbal cues and detection accuracy. *Applied Cognitive Psychology, 25*(2), 212-218. doi:10.1002/acp.1665

Footnotes

[1] There are situations in which true intentions are not by definition high-likelihood events. For example, "I intend to go to bed early all night this week, but I know it is unlikely". Here, we rely on the narrower definition of an intention by Granhag (2010) referring to a single planned action intended to be performed in the near future.

[2] There were deviations from the interview protocol in the Warmelink et al. 2017 study. Occasionally, several questions were merged to one single question, or questions outside the interview protocol were asked (e.g., follow up questions). Interviewees' answers to all questions (i.e., both those in line with and those deviating from the interview protocol) were included for analyses on the study level. However, only answers to interview protocol questions were included for analyses on the question level.

[3] A sensitivity power analysis with alpha level set at .05 revealed that the study had 80% statistical power to detect an effect of Cohen's $d = 0.07$.

[4] The 95% CIs were calculated from the slope +/− (1.96 * standard error). Standard errors were obtained using restricted maximum likelihood estimation. The p-values were calculated based on the Satterthwaite degrees of freedom approximation. The differing assumptions of each of these statistics mean that the p-value can be non-significant even though the confidence interval does not include zero.

[5] As an additional exploratory analysis, we also conducted an analysis on the same corpus of text with English stopwords removed. The results of this analysis were highly similar but suggested a slightly stronger effect of veracity on concreteness, $b = -0.08$, 95% CI [-0.15, -0.02], $t(6.17) = 3.02$, $p = .023$. Unlike the analysis of the full data corpus, this coefficient is significantly different from 0, but the BIC-based Bayes Factor still strongly favors the null hypothesis, $BF_{01} = 37.09$.

[6] All model diagnostics for the multilevel analyses conducted to test H1 and H2 are found in a table in Supplemental material (e.g., random effects).

[7] A full examination of this issue is out of the scope of this paper. Preliminary analyses of the face validity of the folk-concreteness measure and LCM are described in more detail in a blog post (Puddle-Ducks, 2019).

Appendix A

Table A1

Manipulations of true and false intentions used in studies included in the mega-analysis

Study	Description	Manipulation instructions given to true intention condition	Manipulation instructions given to false intention condition
Kleinberg, Nahari, Arntz, & Verschure (2017)	Participants told the truth or lied about a trip	Participants were told to provide honest answers about their trip (e.g., “London for work”). This information was repeated before each interview question.	Participants were assigned a destination (e.g., “Madrid”) and purpose (e.g., “holiday”) and were told to pretend they were planning to fly to this destination with this purpose. This information was repeated before each interview question.
Vrij, Granhag, Mann, and Leal (2011)	Passengers at an airport told the truth or lied about their upcoming trip	Participants were asked to tell the truth about the destination they were flying to. They were first given the following additional instructions (same for truth tellers and liars): “My colleague will ask you a few questions about your forthcoming trip. Some people will be asked to tell the truth whereas others will be asked to lie during these interviews. My colleague, who does not know who is lying or telling the truth, will make a veracity judgement at the end of the interview. When my colleague believes that you are telling the truth, you will get £10, if she thinks that you are lying, you will not get any money.	Participants were asked to lie about the destination they were flying to. They were first given the following additional instructions (same for truth tellers and liars): “My colleague will ask you a few questions about your forthcoming trip. Some people will be asked to tell the truth whereas others will be asked to lie during these interviews. My colleague, who does not know who is lying or telling the truth, will make a veracity judgement at the end of the interview. When my colleague believes that you are telling the truth, you will get £10, if she thinks that you are lying, you will not get any money.

Appendix A

Vrij, Leal, Mann, and Granhag (2011)	Participants honestly or deceptively described their mission	When participants met a friendly agent during their mission, they were instructed to reveal actual details about their mission.	When participants met a hostile agent during their mission, they were instructed to make something up about all aspects of their mission.
Warmelink, Vrij, Mann, & Granhag (2013, Exp. 2)	Passengers on a ferry told the truth or lied about their upcoming trip	Participants were given the following instructions: “An interviewer will come and ask you for details about your trip. Please answer truthfully. Try to convince the interviewer that you are telling the truth.”	Participants were given the following instructions: “An interviewer will come and ask you for details about your trip. Please lie to them and pretend you are making a different trip than the one you in fact are going to make. Try to convince the interviewer that you are telling the truth.”
Warmelink, Vrij, Mann, Jundi, and Granhag (2012)	Participants told the truth or lied about an upcoming trip	Participants were given the following instructions: “In the interview I want you to answer the questions truthfully. Some people are asked to lie during the interview. The interviewer knows that some people may lie, but doesn't know whether you are telling the truth or lying. Your goal is to convince the interviewer that you really are telling the truth.”	Participants were given the following instructions: “In the interview I want you to lie and pretend that you are travelling to [destination] and that you are going there for the purpose of [reason]. The interviewer knows that some people may lie, but does not know whether you are lying or telling the truth. Your goal is to convince the interviewer that you are telling the truth.”
Warmelink (2017)	Participants told the truth or lied about the last day of term	2-3 days before the study the participants were emailed the following information: “You have been placed in the truth telling condition. In order for you to complete the study successfully, could you please email me [address information] what your intentions are for the last day of term ([date]).” Before the	2-3 days before the study the participants were emailed the following information: “You have been placed in the lying condition. In order for you to complete the study successfully, could you please email me [address information] what your intentions are for the last day of term ([date]). In the study I’m going to ask you to lie about this intention. In order to be convincing please prepare a lie. This lie should be plausible

Appendix A

	interview the experimenter asked the participants to confirm that the intention they described in their email was still true. If this was not the case the experimenter offered the participants some time to prepare for the interview. The experimenter reiterated the interview instructions: “The interviewer is trying to detect who is lying and who is telling the truth. You need to convince her that you are telling the truth”	and you should be able to convince somebody you are actually going to do your lie. Please describe that lie […] like you described your real intention” Before the interview the experimenter asked the participants to confirm that the intention they described in their email was still true and that they were still happy that the lie they had created was convincing enough. If this was not the case the experimenter offered the participants some time to prepare for the interview. Then the experimenter reiterated the instructions using the words “The interviewer is trying to detect who is lying and who is telling the truth. You need to convince her that you are telling the truth”