Veracity is in the eye of the beholder:

A lens model examination of consistency and deception.

Charlotte A. Hudson[1], Aldert Vrij[1], Lucy Akehurst[1], Lorraine Hope[1] & Liam P. Satchell[2]

[1]Department of Psychology, University of Portsmouth, UK

[2]Department of Psychology, University of Winchester, UK

Word Count: 6,236

Correspondence concerning this article should be addressed to Charlotte A. Hudson,

Department of Psychology, King Henry Building, King Henry I Street, Portsmouth,

Hampshire, United Kingdom, PO1 2DY. Email: charlotte.hudson@port.ac.uk

Abstract

Word count: 150

Research has attempted to explain perceived cues to deception based upon self-report of what participants believe are 'good' cues to deception, or self-report of what cues participants say they base their veracity judgements on. However, it is not clear to what extent participants can accurately self-report what influences their decision-making. Using a within-subjects design, 285 participants completed a questionnaire regarding their beliefs about deception before rating a selection of truthful and deceptive statements on a variety of cues. Expert coders also rated the statements for the same cues. Laypeople and expert coders do not conceptualise between-subject consistency in the same way. A lens model showed that whilst perceptions of cues, such as consistency and amount of detail, influence veracity judgements, these perceptions (and overall veracity judgements) are mostly inaccurate. Fundamentally, there seems to be inconsistencies between how deception research examines consistency and how it is understood and used by laypeople.

Keywords (5): consistency, deception, veracity judgements, lens model, investigative interviewing

Veracity is in the eye of the beholder:

A lens model examination of consistency and deception.

The "consistency heuristic" is the belief that consistency indicates truthfulness and inconsistency indicates deception (Granhag & Strömwall, 1999). The consistency heuristic is used to inform veracity judgements when examining multiple interviewees' statements (Strömwall, Granhag & Jonsson, 2003), when examining the consistency within statements (Reinhard, Greifeneder & Scharmach, 2013), and when comparing interviewee statements to evidence (Hartwig, Granhag, Strömwall & Vrij, 2005). The consistency heuristic has been mostly examined in relation to multiple statements taken from the same individual (Granhag & Strömwall, 2000; Masip et al., 2018; Strömwall & Granhag, 2003). To date there is little research examining whether people actually *use* the consistency heuristic when making veracity judgements across multiple statements, or whether they simply report it as a perceived cue to deception.

Research has shown the consistency heuristic does not reflect actual truth teller and liar behaviour, as there is little difference between truth teller and liar consistency (Vredeveldt, van Koppen, & Granhag, 2014). As such, consistency is an unreliable cue. In this study, we address three main research questions. First, we examine whether laypeople assess the between-statement consistency of statements in the same way that researchers do. Secondly, we establish whether participants' perceptions of between-statement consistency inform their veracity judgements, and whether these perceptions of between-statement consistency are accurate in classifying statement veracity. Finally, we explore whether participants actually use the cues they report as being important in their veracity decision-making process.

One of the most commonly reported verbal cues to deception is between-statement (in)consistency (Granhag & Strömwall, 2000, 2001); reported by both laypeople (Krix,

Sauerland, Lorei, & Rispens, 2015) and legal professionals (Strömwall & Granhag, 2003).

When examining the perceptions of the deception-consistency relationship amongst

uninformed laypeople and investigators, Masip et al. (2018) found that uninformed laypeople

who were asked to make veracity judgements on a series of written statements most

commonly reported (90%) utilising (in)consistency to assist in making their judgement.

Researchers establish consistency using four main components; repetitions,

omissions, reminiscences, and contradictions (Fisher, Vrij, & Leins, 2013). The presence of

repetitions increase between-statement consistency, whereas omissions, reminiscences and

contradictions decrease between-statement consistency. Yet memory research suggests that

some types of inconsistency, such as reminiscences and omissions, are common features of

memory reports (Fisher, Brewer, & Mitchell, 2009; Gilbert & Fisher, 2006; Strange, Dysart,

& Loftus, 2015).

The general belief in the 'consistency heuristic', as reported by both laypeople and

practitioners, is that between-statement consistency implies honesty and accuracy, and that

any inconsistencies are indicative of deception or inaccurate answers (Strömwall & Granhag,

2003). However, when considering multiple statements from the same individual, the

consistency heuristic is incongruent with both theoretical and empirical memory and

deception research, which to date has shown there is little difference between the consistency

of liars and truth tellers (Vredeveldt, van Koppen, & Granhag, 2014). Even when differences

do occur, the evidence suggests that consistency is more indicative of deception, rather than

honesty (Vredeveldt, van Koppen, & Granhag, 2014).

Whilst there is a large body of research showing that people generally agree how the

consistency heuristic should be used, identifying consistency in a series of consecutive

statements is still idiosyncratic. Granhag and Strömwall (2000) asked participants to act as lie

detectors and rate a series of three videos of the same suspect interrogated on three separate

occasions. When the subjective cues reported as justification for participants' veracity-judgements were examined, consistency was the most commonly reported cue. Of 125 participants, 78 reported using the consistency heuristic in making their veracity judgement. However, 38 of these participants reported that they considered the three consecutive statements consistent over time, whereas the other 40 participants reported that the statements were inconsistent over time. Thus, the exact same series of consecutive statements were considered consistent by one judge, and inconsistent by another. Although people seem to agree that the consistency heuristic is a good cue to deception, they display little agreement as whether a set of statements are consistent or not.

Given these discrepant findings, it seems that whether a series of statements are perceived as consistent or not is moderated by the individual factors pertaining to the judge. It has been suggested that personality traits underpin deception production abilities, and as such should in turn affect lie detection abilities (Semrad, Scott-Parker, & Nagel, 2019). Previous research is contrasting, with lie production skills having been found to facilitate lie detection abilities (Wright, Berry, & Bird, 2012; Wright, Berry, Catmur, & Bird, 2015), and also to have no effect (DePaulo & Rosenthal, 1979; Levine, 2016).

Previous meta-analyses have found no relationship between individual differences and accuracy in judging deception, however the studies included in the meta-analyses did not have adequate sample sizes to explore the relationship between personality dimensions and accuracy in veracity judgements (Aamodt & Custer, 2006). The majority of past research has used samples of $N \leq 75$, however correlations required to determine meaningful relationships stablise at $N > 250$ (Schönbrodt & Perugini, 2013). A recent study with a sample of 207 participants found that dark and maladaptive personality traits (captured in the honesty-humility domain) were associated with judgemental biases, however did not relate to the ability to detect deception (Wissing & Reinhard, 2017). Honesty-humility traits (e.g.

sincerity, fairness) may influence someone to be more likely to consider statements as truthful, whereas social skills may make a person more socially discerning. Therefore there is a need for adequately powered research to examine personality traits and individual lie detection abilities.

To date, researchers have attempted to examine subjective cues to deception based on self-report (see meta-analysis by Hartwig & Bond, 2011). Self-reports usually takes the form of open-ended questions about what participants believe are good cues to deception or what cues they base their veracity judgements on. There is a methodological limitation to this approach, as there is no way to establish that the cues people report using are actually the ones that best explain their decision-making process. It is possible due to low cue validity, people are unaware of what drives their veracity judgements, and when asked about it choose to report explicit social stereotypes about deceptive behaviour.

One approach to address this question it to use lens modelling. The lens model (Brunswik, 1952) was designed to examine the extent to which perception is influenced by differing cues available in a stimulus. Fielder and Walka (1993) used lens modelling techniques to explore the contribution of common cues (e.g. head movements, smiling, speech rate) on judgments of deception. They demonstrated how certain perceived cues predicted lie condition and subjective evaluation when participants were informed about how to interpret the cues and given feedback.

A lens model meta-analysis by Hartwig and Bond (2011) summarised *i)* the relationship between non-verbal and verbal cues created by deception and *ii)* how the non-verbal and verbal cues influenced perceptions of deception, to assess which behaviours were enabling accurate judgments of deception. Through the examination of 81 distinct judgement cues, it was found that individuals often based their veracity judgements on cues that were not indicative of deception. Furthermore, the cues that participants reported using did not

influence their decision-making. Consequently, this meta-analysis suggests that individuals are not only holding false beliefs about valid information for deception detection but are also not aware of their inaccurate beliefs. Similarly, a meta-analysis lens model by Sporer and Schwandt (2007) examined 12 non-verbal cues to deception, taken from 54 studies. Only three of these cues (nodding, foot and leg movements, and hand movements) were found to be reliably associated with deception, whereas those most commonly believed to be associated with deception (adaptors, foot and leg movements, and illustrators) had either opposing or negligible effects, again suggesting that individuals hold false beliefs about valid information for deception detection. However, these meta-analyses do not fit within the traditional lens model approach. For a lens model analysis, all participants should be exposed to the same type of stimuli. In a meta-analytic approach, there is great variability in participants' experiences, stimulus material (e.g. written statements, videos, or live interview) and study methodology.

Research is yet to examine beliefs about the consistency heuristic and veracity judgements using a lens model approach. In the current research, we examined how much of the variance in veracity judgments is explained by essential sources of information present in a pair of statements, such as the level of detail and between-statement consistency. Participants were asked to report their beliefs regarding a variety of cues to deception, before making veracity judgements on a series of statements, and rating them for level of detail and consistency. As not all judges interpret cues equally, and in turn, not all statements contain all cues, using a within subjects repeated measures design we account for the variance that is contributed by the judge (i.e. a participants' personal beliefs about consistency and deception) as well as the statement. Given the widely reported use of between-statement consistency to inform veracity judgements (Granhag & Strömwall, 2001; Masip et al., 2018; Strömwall & Granhag, 2003), it was predicted that participants will use the consistency

heuristic to inform their credibility assessments. As such, participants' perception of between-statement consistency will influence their veracity judgements such that statements that are perceived to be higher in consistency will be rated significantly more truthful than statements that are lower in consistency (Hypothesis 1). As research shows that truth tellers provide more details than liars (DePaulo et al., 2003), it is predicted that statements that are perceived as more detailed will be rated as more truthful than statements that are lower in detail (Hypothesis 2). We also predicted that statements that are perceived to be higher in terms of both between-statement consistency and detail will be considered significantly more truthful than statements that are perceived to be lower in between-statement consistency and lower in detail (Hypothesis 3). It was anticipated that participants will prioritise the consistency heuristic in their veracity decision-making, as it is a more commonly reported perceived cue to deception than amount of detail. As such we predict that statements that are perceived to be higher in between-statement consistency and lower in detail will be considered significantly more truthful than statements that are perceived lower in between-statement consistency and higher in detail (Hypothesis 4). As participants can interpret the same cues in different ways (Granhag & Strömwall, 2000), we predicted there will be variability in veracity ratings. We propose this variability will be greater accounted for by individual factors of the participant rating the statement rather than by level of detail and level of consistency between statements (Hypothesis 5). However, as the effect of individual factors of the participant and the influence of the statement cannot be empirically compared, we report the *SD* of the statement and participant intercepts separately in the model.

Method

Participants

We recruited 285 participants (48 male, 230 female, 6 non-binary, 1 preferred not to say, aged 21-75 years, M_{age} = 40.48 years, SD_{age} = 12.43) to complete an online questionnaire using social media. In total, 772 participants clicked on the survey link, but due to incomplete data (as per pre-registration, completion of the questionnaire was a requirement to include data for analysis, osf.io/vqga4), a final sample of 285 (36.81%) were retained for analysis. Our cessation of data collection rule was that the questionnaire would be taken offline on the Friday after a minimum of 120 participants had completed it. This sample was based upon a priori power calculations indicating that for the study to detect a medium effect size (r = .30) with high power (.90), given alpha = .05, a minimum of 112 participants were required.

Design

This study used a within-subjects design, where participants reported their beliefs about cues to deception as well as their judgments of four statements for perception of: veracity, amount of detail, consistency, number of repetitions, number of omissions, number of reminiscences, number of contradictions, and confidence for their veracity judgement. To explore whether laypeople consider consistency in the same way as expert coders, participants provided an overall rating of between-statement consistency as well as ratings of the four factors of consistency identified by Fisher, Vrij and Leins (2013).

Materials

Beliefs questionnaire. Participants completed a short 20-item questionnaire regarding their beliefs about cues to deception, including six critical questions (regarding detail, consistency, repetitions, omissions, reminiscences and contradictions) and 14 foil questions. The foil questions concerned individuals' beliefs in other identified cues to deception, taken from Reality Monitoring criteria (Sporer & Sharman, 2006), verifiability approach (Nahari, Vrij, & Fisher, 2014) and a meta-analysis regarding cues to deception (DePaulo et al., 2003). We included foil questions to prevent participants identifying that consistency and detail

were the focus of the questionnaire, and do not analyse these beliefs as they are not relevant to our research questions. Participants were asked to rate the cues with respect to how much they believed they were useful cues for determining veracity. Each cue was presented as a statement (for example "Statements that contain unexpected complications") and was rated on an 11-point Likert scale where 0% represented a strong cue to honesty, 50% was neutral, and 100% represented a strong cue to deception. Question order was randomised to reduce order effects, and the beliefs questionnaire took approximately 3 minutes to complete. A copy of the beliefs questionnaire is available at osf.io/ambcf/.

Personality questionnaire. An exploratory component of this study was the inclusion of a personality measure. Participants were asked to complete a short personality questionnaire (HEXACO-60, Ashton & Lee, 2009) as a filler task between reporting their beliefs about cues to deception and making their veracity judgements. The HEXACO-60 is a well-established tool consisting of 60 questions measuring Honesty-Humility (H, here Mcdonald's ω= .72), Emotionality (E, ω= .79), Extraversion (X, ω= .85), Agreeableness (A, ω= .78), Conscientiousness (C, ω= .75), and Openness to Experience (O, ω= .75). The HEXACO-60 questionnaire took approximately 12 minutes to complete.

Rated statements. To generate statements for rating, 17 individuals (6 male, 11 female, aged 18-52, $M_{age} = 28.44$, $SD_{age} = 9.35$) were randomly allocated to either telling the truth or lying about a memorable event that took place in the past two years. They were interviewed by an instruction-blind interviewer. Truth tellers were asked to recall a positive memorable event that occurred within the past two years, and were given examples such as of going on holiday or having a celebratory birthday dinner. They were instructed to choose an event that they were comfortable talking about honestly and could remember in detail. Those who were asked to lie were asked to invent the details of a positive memorable event provided by the researcher (for example "Spending Christmas on a beach"). Following a

delay of one week, all participants returned to be interviewed by the same interviewer for a second time. Participants were either truthful in both interviews or deceptive in both interviews. Interviews were audio recorded and conducted using a standardised script, featuring one open ended request for the interviewee to tell the interviewer everything about their memorable event. The transcripts were then used as stimulus materials within the subsequent online questionnaire.

The memorable event interviews were transcribed and coded for detail and between-statement consistency. Each interview transcript was first coded for number of details provided. (e.g. "on Christmas day itself we went for a walk on the beach which was sandy and freezing cold." would contain five details; 'sandy', 'freezing cold', 'walk', 'Christmas day', and 'beach'). Details were only counted the first time they were mentioned for each account. For both truth tellers and liars, the details provided in their second accounts were compared to those provided in their first accounts, and categorised as the elements of consistency (specified as "repetition", "omission", "reminiscent", and "contradiction", as described by Fisher, Vrij & Leins, 2013). Repetitions were details reported in both phases of the interview, omissions were details reported in the first phase but not in the second phase of the interviews, reminiscent details were details reported in the second phase but not in the first phase of the interviews, and contradictions were details reported in the first phase that were reported differently in the second phase. A subset of four interviews (25%) were coded by a second researcher, who was blind to the experimental conditions. The inter-rater reliability between the coders was high for details in the first accounts (intra-class correlation coefficient [ICC] = .96) and detail in the second accounts (ICC = .98). High reliability was also found across the two coders for repetitions (ICC = .98), reminiscences (ICC = .95), contradictions (ICC = .88) and omissions (ICC = .89). There were no significant differences

between truth teller and liar statements for the amount of detail provided, or any of the factors of consistency, all t's < .81, all p's > .43.

For the online questionnaire, we used the first 16 sets of the paired statements, and disregarded the statements provided by the 17th participant. Participants were shown pairs of statements so that they could compare the details provided at time one to the details at time two to assess between-statement consistency. We chose not to include the 17th pair of statements, to ensure an even number of paired statements were used. This meant that an equal number of truth teller and liar statements could be presented. However, each participant viewed a different combination of four of the paired statements drawn randomly from the pool of 16 paired statements. Thus, all 16 pairs of statements were viewed across the sample, but in different combinations of four per participant, in order to minimise order effects. Participants were required to make one binary veracity decision for each pair of statements they read (truthful or lie) and rate their confidence in their decision on an 11-point Likert scale (where 0% is low confidence and 100% is high confidence). Additionally, the pair of statements were rated on an 11-point Likert scale (where 0% is low, 50% is average, and 100% is high) for the perceived amount of detail, consistency between the statements, number of repetitions, number of omitted details, number of reminiscent details, and number of contradictions. Participants were provided with the definitions for repetitions, omissions, reminiscences, and contradictions (e.g. [Please rate the following statements for] "number of repetitions [i.e. how much information is repeated from statement 1 to statement 2?]", "number of omitted details [i.e. how much information is reported in statement 1 but NOT in statement 2]", "number of reminiscent details [i.e. how much information is reported in statement 2 but NOT in statement 1]", and "number of contradictions [i.e. how much information in one statement opposes information in the other statement?]"). Completing the ratings for four sets of paired statements took on average around 15 minutes to complete.

Procedure

Participants were first presented with the 20-item questionnaire regarding their beliefs about cues to deception. Questions were presented one at a time and in a random order. Following this, participants completed the HEXACO-60 as a filler task, before moving on to the veracity judgement task. Participants were each asked to read a different selection of four randomly selected pairs of statements from the bank of 16 sets of paired statements. Participants were informed that each pair of statements was taken from the same individual one week apart, and that the individual may be lying in both statements or telling the truth in both statements about their experience. Participants were asked to read each pair of statements and select whether they believed the pair of statements were an honest or deceptive report, and provide a confidence rating for their judgement. They were then asked to rate the statement with respect to the amount of detail, consistency between the statements, number of repetitions, number of omitted details, number of reminiscent details, and number of contradictions, which were sequentially presented below the statement pairs, and in a randomised order. Participants were informed that all of the rating variables were of equal importance, and as a statement evaluator, they needed to look for each of the qualities in the statements.

Once the ratings for all four paired statements were completed, participants were debriefed and thanked for their time. Data was collected using online survey platform Qualtrics, and in total took approximately 30 minutes to complete.

Analysis

The statistical tests used in this study (linear mixed effects modelling) as well as the overall analytic framework (lens modelling) and are not widely used in the deception detection literature. However, both offer important insight into the way participants engage with studies like ours and we outline these techniques in more detail here.

Linear mixed effects modelling (using 'lme4' for R; Bates, Mächler, Bolker, & Walker, 2015) is an extension of standard regression techniques for repeated measures data. Whereas standard regression allows one x and one y value per data point, and thus requires aggregating multiple trials or multiple participants for analysis, linear mixed effects modelling can analyse all of the data collected in the study. It does this by computing individual regression intercept and slopes values for so-called 'random' factors. Random factors are quantified sources of potential variance beyond the factors that we have manipulated (which are known as 'fixed factors'). Here, both the 16 statements and the 285 participants are sampled from a wider theoretical total population of possible participants and possible statements and could explain variance in our results. As such the statements and the participants are random factors. Our allocation of statements to being truthful or deceptive is an experimental manipulation and a 'fixed factor' in our design. Linear mixed models, much like standard regression, will allow us to see how participant ratings predict our fixed factor of statement condition and will additionally report how much variance in this effect is explained by our random factors.

We will use these models to evaluate the relationship between participant judgments and statement qualities. So we will firstly investigate the relationship between participant and expert coding of statement features, such as repetitions or omission, whilst controlling for random participant and statement factors. When we report linear mixed effects models in the paper, we report the unstandardised overall sample slope (the 'estimate' in Table 2 or "β = .xx" in text) with 95% CI of that effect ("[.xx, .xx]"), a *p* value and the standard deviation of participant-level slopes ("*SD* = .xx"). An effect with a larger *SD* of slopes suggests more variability in the overall effect at the participant level than an effect with a smaller *SD* of slopes.

Secondly, we will use linear mixed models in a 'lens model' framework to understand the relationship between participant binary judgments of veracity, statement veracity and participant coded cues. A lens model framework allows us to test the 'achievement' of participant veracity judgments at detecting statement veracity and explore how participants' use of the cues they coded *i)* influence their veracity judgment and *ii)* detect the veracity condition of the statements. Lens modelling is distinct to mediation analysis approaches as there is reasonable expectancy that the potential cues are not all correlated with judgments or statement veracity. Whilst it is possible that some cues relate to both participants' judgments and statements (a 'useful cue'), it is also expected that some of the coded cues relate to neither (a 'useless cue'), some coded cues may relate to statement veracity but not judgments (a 'missed cue') or some coded cues may influence judgment but not relate to veracity (a 'red herring'). Lens modelling is often presented graphically to show the effects most efficiently.

Lens models can show which cues participants are using without having to ask participants the overt questions "what criterion influenced your judgment?" as participants may not be aware. By taking advantage of the questionnaires in this study and the extraction of participant-level regression slopes from the linear mixed effects analysis, we can investigate the extent to which a participant reports the importance of a cue on their judgment (i.e. "omissions are 100% a strong cue to deception) and the relationship between the cue and their judgments (i.e. participant 1's omission-judgment relationship was only $\beta = .10$). We can then correlate participant slopes with their self-reported cue usage.

Similarly, correlations will be used to explain the relationship between variance in HEXACO personality and variance in participant-level slopes for the detecting statement veracity with veracity judgments (participant 'accuracy').

Due to the number of tests run in this study, we correct our alpha level so that we only report tests meeting a conservative criteria of $p < .005$ as statistically significant throughout.

All our code for analysis can be found at osf.io/nd8ke/, and data is available at osf.io/a2yxq/ and osf.io/39qby/.

Results

Expert-Participant Agreement on Cues.

To establish whether laypeople interpreted statements in the same way expert coders are trained to, we examined a linear mixed effects model (Table 1). We found that participants' ratings of details concurred with those given by expert coders for the first account and the second account. To a lesser extent, participants agreed with expert coders about the number of reminiscences. There was no such agreement between participants and expert coders for omissions, contradictions, or repetitions.

[Insert Table 1 about here]

Veracity cue perception and judgements.

We explored whether participants' perceptions of between-statement consistency influenced their veracity judgements, and whether these perceptions of consistency were accurate in identifying truthful or deceptive statements. Participants were accurate in identifying statement veracity 50.88% of the time, and accurate in identifying truthful statements 65.63% of the time. First we discuss the relationship between participants' ratings of the statements and their binary judgment of veracity. Full linear mixed effects models are reported in Table 2. We found that when participants rated statements as more detailed, they were more likely to consider statements as truthful, supporting Hypothesis 1. Furthermore, when participants rated statements as more consistent, they were significantly more likely to consider statements as truthful (see Table 2), supporting Hypothesis 2. In addition, fewer contradictions, fewer omissions, and more repetitions were significantly more likely to influence veracity perceptions to be truthful (all $p < .001$, see Table 2 and Figure 1). The relationship between participant ratings of reminiscence and veracity judgments did not meet

our adjusted alpha level ($p < .005$) as $p = .019$. No interaction was found between consistency and detail, ($\beta = 0.20$ [-0.10, 0.42], $p = .062$, $SD = 0.41$), and as such no support was found for Hypotheses 3 and 4.

On the other side of the lens (see left of Figure 1, Table 2), we investigated how participant ratings of cues could be used to predict the veracity of the statements. Higher participant ratings of the category of consistency did predict honest statements, $p < .001$, however none of their judgements of the four factors of consistency and the amount of detail were predictive of the veracity condition, all p's $\geq .937$.

The participants' veracity judgments did not predict statement veracity ($\beta = 0.03$ [-0.96, 1.03], $p = .948$, $SD = 0.32$), with the SD intercept for participant influence $= 0.09$, and the SD for the statement influence $= .47$. However, the effect of the statement was notably heterogenous ($SD = 0.47$) suggesting that there were statements which participants were reasonably accurate about, and statements that individuals were reasonably inaccurate about. The effect of participants was smaller ($SD = 0.09$), and so no support is found for Hypothesis 5.

[Insert Figure 1 about here]

[Insert Table 2 about here]

Actual and self-reported cue usage.

We next investigated the relationship between self-reported cue influence and the extracted participant-level slopes of cue usage (i.e. their actual use of the cue, see the right-hand side of Figure 1) to explore whether participants were utilising the cues they report as being important to the decision-making process. A significant negative correlation was found between the participants self-reported utility of cues, and participants actual use of cues for reminiscence, $r(283) = -.17$, [-.38, -.06], $p = .003$. No correlation was found between self-reported utility and actual use for amount of detail, $r(283) = -.03$, [-.15, .09], $p = .612$,

consistency, $r(283) = -.13$, [-.24, -.02], $p = .026$, contradictions, $r(283) = -.04$, [-.16, .08], $p =$ .504, omissions, $r(283) = .06$ [-.05, .18], $p = .283$, or repetitions, $r(283) = -.04$ [-.15, .07], $p =$ .470. In other words, participants did not show insight into which features of statements they were using to form veracity judgments.

We also examined correlations between participants' holistic consistency judgements, and their ratings of the four components of consistency. A strong positive correlation was found between the perception of consistency and the perceived number of repetitions, $r = .64$ [.61, .68], $p < .001$. We also found a strong negative correlation between the perception of consistency and the perceived number of contradictions, $r = -.73$ [-.75, -.70], $p < .001$, and a moderate negative correlation between the perception of consistency and the perceived number of omissions, $r = -.46$ [-.50, -.41], $p < .001$. Finally, a weak correlation was found between the perception of consistency and the perceived number of reminiscent detail, $r = -.20$ [-.25, -.14], $p < .001$.

Personality of perception of deception.

We conducted exploratory analyses to look at participant HEXACO personality traits and accuracy of veracity judgements to identify whether there was a difference between the personality traits of those who were accurate and inaccurate in their veracity judgements. There were no notable correlations between participant personality traits and accuracy of veracity judgements (all absolute $r \leq .14$, all $p \geq .02$).

Discussion

We examined the relationship between participants' perceptions of between-statement consistency, their perceptions of four components of consistency (repetitions, omissions, reminiscences, and contradictions), and their veracity judgements. This research addressed three main questions. First, we wanted to establish whether laypeople would interpret statements in the same way as trained expert coders. The four components of consistency

were operationalised by expert coders, and the participants' rating of the components were compared to this. We operationalised between-statement consistency using expert coders, as this is how consistency is commonly determined and empirically tested within a research setting. We found that while participants reported perceiving the same amount of overall detail and reminiscent detail in a statement as the expert coders, they did not agree on the amount of omission, repetition, or contradiction. This would suggest that expert coders and laypeople are not relying on the same cues when they assess between-statement consistency across multiple statements. Therefore, research reporting the behaviours of experts who code statements for between-statement consistency in detail, with dedicated time and resources, may not examine the same behaviour as novices trying to identify 'consistency' for the first time. In this research, participants' holistic consistency judgements accurately predicted the veracity condition of a statement, but did not seem to relate to the components of consistency. It is therefore important for future research to more precisely delineate what is meant by participants when they report utilising consistency to inform their veracity judgements.

Secondly, we explored whether participants' perceptions of between-statement consistency influenced their veracity judgements, and whether these perceptions of consistency were accurate in identifying truthful or deceptive statements. We found that perception of amount of detail, holistic consistency, contradictions, omissions, and repetitions all significantly contributed to the veracity judgement, however the perception of reminiscence did not. Greater holistic consistency increased the likelihood of truthful judgements (although there was a large amount of variance in how this cue was interpreted), as would be suggested by the consistency heuristic. In line with the consistency heuristic and the subcomponents of consistency identified by Fisher, Vrij and Leins (2013), greater amounts of repetition (an increase in consistency) led to more truthful judgements, and was positively correlated with participants' holistic perceptions of consistency. Conversely,

greater amounts of omission and contradiction (a decrease in consistency) led to more judgements of deception, and both components were negatively correlated with participants' holistic perceptions of consistency. Reminiscence was not used in the formation of veracity judgements and had the weakest correlation with holistic consistency, which may be due to reminiscence being prevalent across truth tellers repeated accounts (Gilbert & Fisher, 2006).

Participants did not make accurate veracity judgements about the statements, with an average accuracy rate of 50.88%, however there was a large amount of variance at the statement-level. This variance indicates that whilst participants may have been reliably accurate or inaccurate in their veracity judgements across certain statements (truthful statements were accurately identified 65.63% of the time), there were also statements where participants judgements varied, suggesting that perception of more ambiguous statements differ from person to person. No notable correlations were found between individuals' personalities and their judgement accuracy. These findings are in line with Wissing and Reinhard (2017), who found no relationship between dark and maladaptive personality traits and lie detection accuracy. Future research would benefit from establishing which individual differences influence veracity judgements, and how accurate judges differ from inaccurate judges.

Finally, we were interested in whether participants used the cues they reported as being important to the decision-making process. There is a possibility that extraneous cues that were not examined informed participants' veracity judgements, and that the cues examined in the current research simply correlate with these extraneous cues. We therefore interpret the following results with caution. A significant negative correlation was found between participants' self-reported use and actual use of reminiscence, however no other correlations between self-reported use and actual use were found. This suggests that although participants reported using reminiscence in their judgement-making, they displayed no

evidence of this when completing their statement ratings. It is possible that the reminiscence could simply be a correlate of deception judgements, as opposed to being the cause of the deception judgements (despite it being self-reported as useful). Regardless, this discrepancy suggests a poor level of introspection for identifying the features of the statements used to make veracity judgements, and therefore a lack of insight regarding intended cue-usage. It is likely that individuals report a priori, implicit causal theories for their behaviours (Nisbett & Wilson, 1977), and therefore self-reported cue usage does not represent participants actual cue usage. We opted to use real statements, in which the consistency was not experimenter-manipulated in order to maximise ecological validity. Future research could benefit from using manufactured statements, where the four components of consistency are deliberately manipulated and all other features are held constant, to determine whether between-statement consistency is used in veracity judgement-making.

There are a number of limitations to the current study, which may have implications for the results. The online questionnaire took around 30 minutes to finish, with only 37% of those who initially clicked the survey link completing all questions. It is reasonable to expect a level of fatigue and a biased sample for a voluntary online study of this length. Participants were asked to make multiple ratings of the same statements on different cues, which may have lowered the attentiveness of participants and resulted in less discrimination in cue perception. Statement presentation was randomised to statistically correct for this, however the repeated nature of the task may have influenced subsequent judgements, and therefore there it a chance that cue perception altered across the task as a function of fatigue. Future research could examine a detection-accuracy curve when participants make multiple judgements, to establish whether previous judgements influence future judgements.

Participants were asked to make binary "truthful" or "deceptive" judgements about statements. However, it is possible that some participants believed the statements contained

embedded lies, and as such were neither fully truthful nor deceptive, but were forced to make a choice. Participants may have made different binary judgements based upon their own biases. This study examined individual-level performance, and consequently these biases would have been accounted for in the participants' random slopes although it cannot be extrapolated from this dataset where these biases may have influenced the decision-making process.

A large amount of deception occurs online over a number of interactions in written format, such as internet scams, financial fraud, and online grooming. In this research, participants were asked to make third-party veracity judgements based upon a written transcript of an interaction. It stands to reason that the scale labels "low amount", "average amount", and "high amount" are somewhat ambiguous and could be interpreted differently without the social context of being involved in the interaction. For instance, what is considered an average amount of detail in a statement may vary, dependent on the sender and the context of the situation. Therefore, to establish what the "average amount" of anything is, based on a written transcript provided by a stranger with little surrounding context to the interaction, is a challenging task, and can be considered a limitation of the design.

In conclusion, this research highlights several key points. Primarily, there seems to be a discrepancy between how research explores between-statement consistency and how it is understood by laypeople. Further to this, no notable correlations were found between self-reported reliance on cues and actual usage of cues, and as such it appears that individuals do not understand the cues they personally use to make veracity decisions. Further exploration is needed to establish what cues reliably influence veracity perceptions, and how individuals vary in the decision-making process. The lens model highlights that perceptions of factors of consistency and amount of detail directly influence the veracity judgement, however these perceptions are often inaccurate, and consequently, so are the veracity judgements. A holistic

perception of consistency could facilitate accurate perceptions of veracity, however it does

not appear to be conceptualised in the same way it is constructed by researchers.

References

<author names redacted>. (2019). *A lens model examination of consistency and deception.*

[data file]. Open Science Framework. DOI: 10.17605/OSF.IO/3DWUB

Aamodt, M. G., & Custer, H. (2006). Who can best catch a liar?. *Forensic Examiner, 15*(1),

6-11.

Ashton, M. C., & Lee, K. (2009). The HEXACO–60: A short measure of the major

dimensions of personality. *Journal of Personality Assessment*, *91*, 340-345. DOI:

10.1080/00223890902935878

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects

Models Using lme4. *Journal of Statistical Software, 67*, 1 - 48. DOI:

10.18637/jss.v067.i01

Brunswik, E. (1952). *The conceptual framework of psychology.* Chicago, IL: University of

Chicago 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. DOI: 10.1037/0033-

2909.129.1.74

DePaulo, B. M., & Rosenthal, R. (1979). Telling lies. *Journal of Personality and Social

Psychology, 37*, 1713-1722. DOI: 10.1037/0022-3514.37.10.1713

Fiedler, K., & Walka, I. (1993). Training lie detectors to use nonverbal cues instead of global

heuristics. *Human Communication Research, 20*(2), 199-223.

Fisher, R. P., Brewer, N., & Mitchell, G. (2009). The relation between consistency and

accuracy of eyewitness testimony: Legal versus cognitive explanations. In T.

Williamson, R. Bull, & T. Valentine (Eds.)*, Handbook of psychology of investigative

interviewing: Current developments and future directions* (pp. 121-136). Chichester:

Wiley-Blackwell

Fisher, R. P., Vrij, A., & Leins, D. A. (2013*). Inconsistency as a predictor of memory inaccuracy and lying. Applied issues in investigative interviewing, eyewitness memory, and credibility assessment.* New York, NY: Springer.

Gilbert, J. A., & Fisher, R. P. (2006). The effects of varied retrieval cues on reminiscence in eyewitness memory. *Applied Cognitive Psychology*, *20*, 723-739. DOI: 10.1002/acp.1232.

Granhag, P. A., & Strömwall, L. A. (1999) Repeated interrogations–stretching the deception detection paradigm. *Expert Evidence*, *7*, 163–174. DOI: 10.1023/A:1008993326434.

Granhag, P. A., & Strömwall, L. A. (2000). Effects of preconceptions on deception detection and new answers to why lie-catchers often fail. *Psychology, Crime and Law, 6*, 197-218. DOI: 10.1080/10683160008409804

Granhag, P. A., & Strömwall, L. A. (2001). Deception detection: Interrogators' and observers' decoding of consecutive statements. *The Journal of Psychology*, *135*, 603-620. DOI: 10.1080/00223980109603723.

Hartwig, M., & Bond Jr, C. F. (2011). Why do lie-catchers fail? A lens model meta-analysis of human lie judgments. *Psychological Bulletin, 137*(4), 643-659. DOI: 10.1037/a0023589

Hartwig, M., Granhag, P. A., Strömwall, L. A., & Vrij, A. (2005). Detecting deception via strategic disclosure of evidence. *Law and Human Behavior, 29*, 469-484. DOI: 10.1007/s10979-005-5521-x

Krix, A. C., Sauerland, M., Lorei, C., & Rispens, I. (2015). Consistency across repeated eyewitness interviews: Contrasting police detectives' beliefs with actual eyewitness performance. *PloS one*, *10*, e0118641. DOI: 10.1371/journal.pone.0118641.

Levine, T. R. (2016). Examining sender and judge variability in honesty assessments and deception detection accuracy: Evidence for a transparent liar but no evidence of

deception-general ability. *Communication Research Reports, 33*, 188-194. DOI: 10.1080/08824096.2016.1186629

Masip, J., Martínez, C., Blandón-Gitlin, I., Sánchez, N., Herrero, C., & Ibabe, I. (2018). Learning to detect deception from evasive answers and inconsistencies across repeated interviews: A study with lay respondents and police officers. *Frontiers in Psychology, 8*, 2207. DOI: 10.3389/fpsyg.2017.02207

Nahari, G., Vrij, A., & Fisher, R. P. (2014). Exploiting liars' verbal strategies by examining the verifiability of details. *Legal and Criminological Psychology, 19*, 227-239. DOI: 10.1111/j.2044-8333.2012.02069.x

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review, 84*(3), 231-259.

Reinhard, M. A., Greifeneder, R., & Scharmach, M. (2013). Unconscious processes improve lie detection. *Journal of Personality and Social Psychology, 105*, 721-739. doi: 10.1037/a0034352

Schönbrodt, F. D., & Perugini, M. [2013]. At what sample size do correlations stabilize? *Journal of Research in Personality, 47*, 609-612. DOI: 10.1016/j.jrp.2013.05.009

Semrad, M., Scott-Parker, B., & Nagel, M. (2019). Personality traits of a good liar: A systematic review of the literature. *Personality and Individual Differences, 147*, 306-316. DOI: 10.1016/j.paid.2019.05.007

Sporer, S. L., & Sharman, S. J. (2006). Should I believe this? Reality monitoring of accounts of self-experienced and invented recent and distant autobiographical events. *Applied Cognitive Psychology, 20*, 837-854. DOI: 10.1002/acp.1234

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

Strange, D., Dysart, J., & Loftus, E. F. (2015). Why errors in alibis are not necessarily

evidence of guilt. *Zeitschrift für Psychologie*. *222,* 82-89. DOI: 10.1027/2151-

2604/a000169.

Strömwall, L., & Granhag, P. A. (2003). How to detect deception? Arresting the beliefs of

police officers, prosecutors and judges. *Psychology, Crime and Law*, *9*, 19-36.

DOI: 10.1080/10683160308138.

Strömwall, L. A., Granhag, P. A., & Jonsson, A. C. (2003). Deception among pairs: "Let's

say we had lunch and hope they will swallow it!". *Psychology, Crime & Law, 9*, 109-

124. DOI: 10.1080/1068316031000116238

Vredeveldt, A., van Koppen, P. J., & Granhag, P. A. (2014). The inconsistent suspect: A

systematic review of different types of consistency in truth tellers and liars.

In *Investigative Interviewing* (pp. 183-207). New York: Springer.

Wissing, B. G., & Reinhard, M. A. (2017). The dark triad and the PID-5 maladaptive

personality traits: accuracy, confidence and response bias in judgments of veracity.

Frontiers in Psychology, 8, 1549. DOI: 10.3389/fpsyg.2017.01549

Wright, G. R., Berry, C. J., & Bird, G. (2012). "You can't kid a kidder": association between

production and detection of deception in an interactive deception task. *Frontiers in

Human Neuroscience, 6*, 87. DOI: 10.3389/fnhum.2012.00087

Wright, G. R., Berry, C. J., Catmur, C., & Bird, G. (2015). Good liars are neither 'dark'nor

self-deceptive. *PloS One, 10*(6), e0127315. DOI: 10.1371/journal.pone.0127315

Table 1. *Expert-participant agreement on cues.*

	Estimate [95% CI]	*p*	*SD*
First account	0.51 [0.34, 0.68]	<.001*	0.01
Second account	0.54 [0.40, 0.69]	<.001*	0.02
Reminiscences	0.34 [0.16, 0.52]	.002*	0.17
Omissions	0.25 [0.01, 0.50]	.063	0.04
Contradictions	0.32 [0.11, 0.53]	.009	0.05
Repetitions	0.21 [0.01, 0.43]	.088	0.01

	Estimate [95% CI]	*p*	*SD*

Table 2. *Full report of linear mixed effects models.*

Cue	Estimate [95% CI]	*p*	Random effect of participant ratings	Random effect of participant slopes	Random effect of statement effects
Cue effects predicting rater veracity *judgements* (Truth coded as +0.5 and Deception coded as -0.5)					
Details	0.55 [0.35, 0.74]	<.001	0.00	0.54	0.93
Holistic Consistency	2.47 [1.97, 2.96]	<.001	0.66	1.39	0.51
Contradictions	-1.49 [-1.78, -1.20]	<.001	0.48	0.48	0.61
Omissions	-0.60 [-0.80, -0.41]	<.001	0.24	0.30	0.83
Repetitions	1.25 [0.95, 1.54]	<.001	0.25	0.85	0.74
Reminiscences	-0.21 [-0.38, -0.04]	.019	0.27	0.49	1.03
Cue effects predicting statement veracity *condition* (Truth coded as +0.5 and Deception coded as -0.5)					
Details	-0.02 [-3.87, 3.82]	.991			95.87
Holistic Consistency	0.11 [0.11, 0.12]	<.001			101.10
Contradictions	-0.10 [-3.88, 3.67]	.958			95.81
Omissions	-0.01 [-3.77, 3.76]	.997			95.87
Repetitions	0.12 [-3.54, 3.78]	.949			95.79
Reminiscences	-0.15 [-3.84, 3.55]	.937			95.74
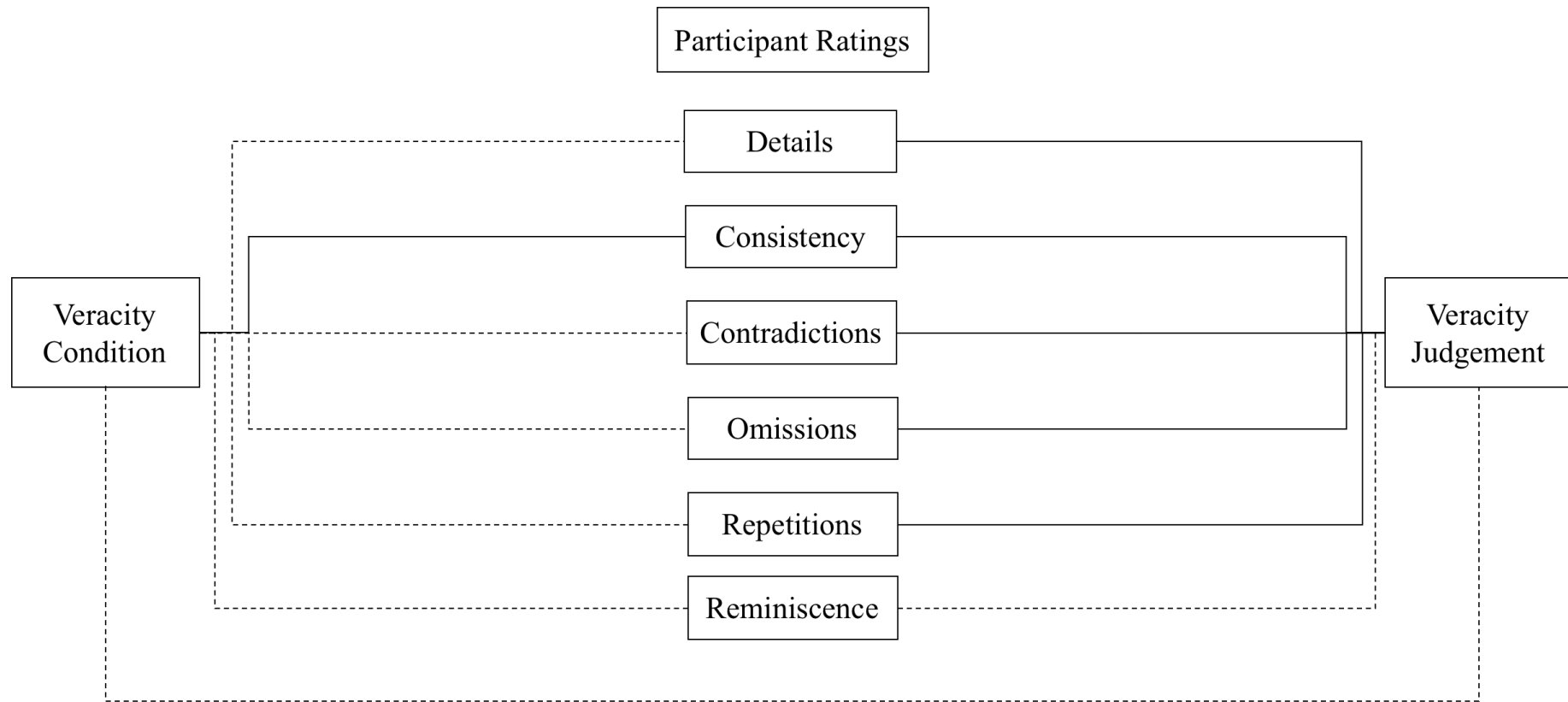
Figure 1. Lens model representing the effects of participant ratings of cues on veracity judgement (right hand side) and veracity condition (left hand side). Bold lines signify $p < .001$, dotted lines are not significant. See Table 1 for values.