The Source of the Truth Bias: Heuristic Processing?

Chris N. H. Street

University of British Columbia

Jaume Masip

University of Salamanca

In press. *Scandinavian Journal of Psychology*.

Author Note

Chris N. H. Street, Department of Psychology, University of British Columbia, Canada; Jaume Masip, Department of Social Psychology and Anthropology, University of Salamanca, Spain.

We are grateful to Jesús López and Margarita Hernández for their help in setting up and letting us use a room in the CRAI to collect the data of Experiments 3 and 4. We would also like to thank Maarten Speekenbrink and Richard Morey for their early discussions on statistical approaches.

Correspondence concerning this article should be addressed to Chris N. H. Street, Department of Psychology, University of British Columbia, 2136 West Mall, Vancouver, BC, V6T 1Z4, Canada. E-mail: c.street@psych.ubc.ca

Word Count: 7,394

Abstract

People believe others are telling the truth more often than they actually are, called the truth bias. Surprisingly, when a speaker is judged at multiple points across their statement the truth bias declines. Previous claims argue this is evidence of a shift from (biased) heuristic processing to (reasoned) analytical processing. In four experiments we contrast the heuristic-analytic model (HAM) with alternative accounts. In Experiment 1, the decrease in truth responding was not the result of speakers appearing more deceptive, but was instead attributable to the rater's processing style. Yet contrary to HAMs, across three experiments we found the decline in bias was not related to the amount of processing time available (Experiment 1-3) or the communication channel (Experiment 2). In Experiment 4 we find support for a new account: that the bias reflects whether raters perceive the statement to be internally consistent.

Keywords: dual-process theory; deception detection; truth bias; heuristic processing; consistency; smart lie detector.

Introduction

When judging if someone is lying or not, naïve observers are biased towards believing the speaker is telling the truth (Bond & DePaulo, 2006; Levine, Park, & McCornack, 1999; Vrij, 2008). A number of explanations have been proposed for this *truth bias*, such as the availability, anchoring, or falsifiability heuristics, or social conversational rules and self-presentational concerns of the deception judge (see Vrij, 2008). Among relational partners, the truth bias has been attributed to relational trust (see Miller and Stiff, 1993). According to Bond and DePaulo's (2006) double standard framework, people believe liars are tormented, shameful, and conscience stricken, and so display nervous behavior. Since lies told in laboratory settings are low-stake and most everyday lies are of little consequence and easy to rationalize by the liar, liars normally do not show indicators of anxiety, shame or guilt, and hence are judged to be honest (Bond & DePaulo, 2006). Our starting point here is still another explanation of the truth bias, one that contends that it is a consequence of heuristic processing.

Humans can process information either heuristically or analytically. Heuristic processing results in fast, intuitive judgments (Evans, 2007) and consumes little cognitive resources (Chaiken, Liberman & Eagly, 1989). Unfortunately, it shows systematic biases (Chaiken et al., 1989). Analytic processing is slower and requires greater effort and cognitive resources than heuristic processing, but it may be less biased (Chaiken et al., 1989).

Although there are diverse characterizations of heuristic-analytic models (HAMs; see, for instance, Chaiken & Trope, 1999; Gilbert, 1999), we used Evans's (2007) taxonomy because it allows for testable predictions to be generated from a broad and general theoretical framework (see Evans & Stanovich, 2013). Other authors, however, have come to similar classifications (e.g., Gilbert's, 1999, corrective, competitive, and selective designs correspond closely to the three kinds of HAMs proposed by Evans). Evans (2007) identified several general classes of HAMs. The first class, called *default-interventionist models*, claims that heuristic processing is the default processing mode, but it can be interrupted by analytic processing provided enough time is available. The second class, *parallel-competition models*, proposes both heuristic and analytical processes run simultaneously. However, if only a short processing time is available heuristic processing is more likely to be the basis of a judgment. Crucially, in both cases *heuristic processing is more likely with shorter processing times* (Evans, 2007; Kahneman & Frederick, 2002; Stupple & Ball, 2008; see also Thompson, Striemer, Reikoff, Gunter & Campbell, 2003; Trippas, Verde & Handley, 2014).

Unlike analytical processing, heuristic processing might yield truth-biased judgments. According to Gilbert, Krull and Malone (1990), incoming information (e.g., a communication message) is first "represented as true before their validity can be rationally assessed" (p. 611), and "disbelief requires extra effort" (Vrij, 2008, p. 149). While heuristic processing is fast, automatic and immediate, the extra-effort needed to analytically assess a message to see whether it should be disbelieved requires time. Therefore, if processing time is short, then truth judgments will be more likely than if processing time is long.

Research has supported this notion. Interrupting people while they are deciding if a smile is true or false creates a truth bias (Gilbert et al., 1990, Study 2), and time pressure similarly increases the likelihood of believing information (Gilbert, Tafarodi & Malone, 1993). Also, to analytically assess a message enough information must be available; this information may be absent if the message is too short. Because of these reasons, the truth bias should be more pronounced with shorter processing times.

Typically, in deception detection research, videotaped truths and lies are shown to observers who have to immediately judge whether each sender in the videotape is lying or telling the truth. Masip, Garrido, and Herrero (2006, 2009, 2010) observed that in many studies the video clips are so brief and potentially uninformative that analytical processing may not happen. They reasoned that this might explain why deception research has typically found a truth bias. Masip et al. conducted two studies in which people watched a video recorded mock crime and then lied/told the truth in answering three questions about the crime. Their answers were videotaped and subsequently shown to raters. The raters judged each speaker's honesty after each of the three responses, aware that any given speaker either always lied or always told the truth across the three responses of their statement. Consistent with a heuristic account, raters were truth biased when judging the speaker's first response, but became less biased when judging the second and third responses.

Although Masip et al. (2010) acknowledged that their results are open to alternative interpretations, they favoured the HAM interpretation. Here we first considered whether the change in bias could be explained simply as a change in the speaker's behaviour, rather than having anything to do with the cognitions of the rater. Having found support for a cognitive account of the bias (Experiment 1), we then moved on to more closely examine two competing cognitive accounts, namely HAM-based (Experiments 1-3) and step-by-step (Experiments 3-4) response mode explanations.

Account 1: Behavioural Explanation

Labelling excessive truth responding as a "bias" implies it is an erroneous tendency of the rater. Yet the truth "bias" may not be a cognitive bias, but a valid inference made from the available behaviours. That is, it is not the raters who become less truth-biased, but rather the speakers who give off cues that appear less honest. For instance, liars may conceivably become more nervous over time or display suspicious behaviours. Truth-tellers, generally confident that “the truth will come out” (Kassin, 2005; Masip & Herrero, 2013), may not. In this case, truth-tellers’ behaviours would appear honest across their statement whereas liars would become increasingly unconvincing. This could explain the phenomenon of the truth bias declining over time.

To begin to support a heuristic-analytical account, we had to show that the decline in truth bias occurs independently of the senders’ behaviour, i.e., that there is a cognitive component to the bias. We adopted Masip et al.’s (2009) paradigm. Observers watched video-recorded speakers giving three consecutive truthful or deceptive responses to an interviewer’s questions. Observers had to indicate whether each speaker was lying or telling the truth after watching each of the three responses. It is important to note the presentation order (order in which each speaker’s responses were shown to raters) in Masip et al.’s studies was the same as the recording order (order in which the responses had been recorded). We reversed the presentation order in one condition so that the last *recorded* response was viewed first and vice versa. The behavioural account predicts that in the reverse order the truth bias should *increase* over successive judgments because speakers would appear more honest in their first recorded (last presented) response. The cognitive account predicts that the bias should decline irrespective of condition.

Account 2: Heuristic-Analytic Account

Masip et al. (2006, 2009) found a decline in the truth bias between each of three speaker’s responses. Although the judgment number (1st, 2nd, 3rd) is a proxy of viewing time, some speakers provided lengthy responses whereas others were shorter, and so it is an inaccurate proxy. Analytic processing should intervene at a given *time*, not after a given number of ratings: the analytical process either takes longer to engage (default-interventionist models) or longer to complete processing (parallel-competitive models) (Evans, 2007). For this reason, we examined the following two issues:

(a) How the total amount of processing time available (i.e., the cumulative speaking time until the end of the third response) influenced bias and accuracy.

(b) How the duration of the first presented response influenced bias and accuracy. This was examined because if the speaker's initial response is particularly long, raters may shift to analytical processing during the first response.

To support a HAM account, the truth bias should decrease and accuracy should increase when there is greater time to process information.

Account 3: Step-by-Step Response Mode

If processing time cannot explain the decline in bias, but judgment number can, then it must be the very act of making multiple judgments that has a causal effect in the decision process. Granhag and Strömwall (2001a) showed that raters who made a judgment after watching an interview and another judgment after watching two subsequent interviews were more accurate than those who only made a single judgment after watching all three interviews. Importantly, in these two conditions the total viewing time was identical. The authors explained the improved accuracy in terms of assessment through a *step-by-step response mode* (making repeated assessments of veracity) instead of an *end-of-sequence response mode* (making just one final assessment). By reflecting on previous decisions in light of new information, observers using a step-by-step response mode could attain greater accuracy.

By examining the impact of viewing duration on truth judgments and accuracy, we also sought to test the influence of judgment time separately from the act of making a judgment. To support a step-by-step response mode account, a decline in biased responding should be evident across the multiple ratings, but it should be independent of the time the rater had to process the information.

In summary, we addressed three potential explanations for the decline in truth judgments across a speaker's statement. The first explanation proposes it may be that speakers appear more deceptive over time (the behavioural account), meaning the bias shift is attributable to speakers' actions, not to the raters' processing style. If there is a cognitive component to the bias, there are two additional explanations. The second, currently favoured explanation is that the truth bias reflects a shift from heuristic to analytical processing (the HAM account). In this case, the decline should be related to the amount of processing time available. The third explanation suggests it is the act of making multiple judgments that causes the decline in bias (the step-by-step account).

EXPERIMENT 1

In line with Masip et al. (2006, 2009, 2010) and with a cognitive account of the bias, we predicted:

(a) That *the proportion of truth judgments* (PJT) would decline over successive ratings irrespective of whether the speaker's first or last recorded response was presented initially.

(b) That *accuracy* would improve over successive ratings.

Because using the point of judgment may not be a valid proxy of time, we also examined:

(c) The *cumulative duration* of the speaker's responses until the moment of rating.

(d) The influence of the *duration of the speaker's first presented response* on raters' PJT and accuracy.

In both cases, longer durations were expected to yield a smaller PJT and greater accuracy rates.

Method

Eighty-three psychology undergraduates (66 female; age M = 20.75, SD = 0.20, range: 18 - 35) participated.

Materials

The video stimulus set was adopted from Masip et al. (2006, 2009). Speakers (n = 24) were shown one of two videotapes with a scene depicted by three characters. The videotapes displayed either an attempted or successfully completed theft. After viewing the footage, speakers were interviewed twice about the actions of each character in the videotape. They had to respond honestly during one interview and deceptively during the other. Both interviews had the same questions, which were: “Describe in detail what the man with a moustache [man in a suit/woman] did; I remind you that you have to tell the truth [lie]”. Question order, whether the speaker lied/told the truth first, and mock-crime videotape were counterbalanced. Later, the 48 video-recorded interviews (24 truthful and 24 deceptive) were divided into four video sets with six liars and six truth tellers in each set. The same speaker never appeared lying and telling the truth in the same set. For the current study, we selected the video set that achieved PJT and accuracy ratings most representative of the overall results in Masip et al.’s (2009) research (Video Set A1). Further details can be found in Masip et al. (2006). The first recorded response lasted on average 50 s, Response 2 averaged 37 s and Response 3 averaged 39 s. Duration differences were not statistically significant, F (2, 35) = 0.26, p = .776, η_p^2 = 0.01.

Two versions of Video Set A1 were created. In the first version (used in the *direct viewing condition*) the three responses of each speaker were presented in the same order in which they had been recorded. In the second version (used in the *reverse viewing condition*) the speaker's third recorded response was presented first, followed by the second recorded response, and then by the first one.

Design and Procedure

Participants were allocated randomly to the direct (n = 44) or reverse viewing condition (n = 39). Sex and age distributions did not differ substantially between conditions. The procedure replicated Masip et al.'s (2009), with the exception of the viewing direction manipulation.

Groups of participants took part in two sessions per condition. They were apart so they could not see each other's responses. Instructions explained they would see 12 speakers, that each speaker provided a single statement based on a videotaped event, that each statement consisted of three responses, and that a statement was either deceptive or truthful across the three responses. After each response, the video was stopped and the participants marked in a booklet their lie-truth judgment and their confidence (on a 1-to-7 scale, with higher values indicating more confidence). Raters were explicitly told that in judging each speaker's 2nd or 3rd response they were free to either make the same judgment or to change it if they changed their opinion. In analysing the data we examined the changes in the raters' judgments across the three responses of each sender, not across different senders, statements, or topics. In other words, the primary independent variable was the speaker's response (1st, 2nd, 3rd), not statement number (1 through 12).

The PJT and accuracy scores were the dependent variables. These measures are typically used in lie detection research. But because they share variance, we also used nonparametric signal detection measures B''_D and A' to measure the effects of response bias and accuracy independently of each other (see the Appendix for more information).

Greenhouse Geisser corrections were used in all instances where assumptions of sphericity were violated.

Results

Testing the Behavioural Account

Two 2 (Veracity: truthful/deceptive statement) x 3 (Presented Response: 1st/2nd/3rd) x 2 (Viewing Direction: direct/reverse) Analyses of Variance (ANOVAs) with repeated measures on the first two variables were conducted. The first ANOVA on the PJT revealed truthful statements were more often judged truthful ($M = .62$, $SD = .02$) than deceptive statements ($M = .53$, $SD = .02$), $F(1, 81) = 16.55$, $p < .001$, $\eta_p^2 = 0.17$. A significant main effect of presented response, $F(1.62, 131.58) = 21.20$, $p < .001$, $\eta_p^2 = 0.21$, reflected a decrease in truth judgments over time that was significant between the first ($M = .61$, $SD = .16$) and the second ($M = .56$, $SD = .16$) and between the first and the third ($M = .54$, $SD = .17$) response of the speakers ($t(82) = 3.48$, $p < .001$, $d = 0.40$, and $t(82) = 4.08$, $p < .001$, $d = 0.46$, respectively), but not between the second and third responses, $t(82) = 0.60$, $p = .155$, $d = 0.08$. A linear contrast analysis found a linear effect of presented response, $F(1, 246) = 8.88$, $p = .003$, $\eta_p^2 = 0.04$, but no quadratic effect, $F(1, 246) = 1.26$, $p = .262$, $\eta_p^2 = 0.01$. The Presented Response x Viewing Direction interaction was not significant, $F(1.62, 131.58) = 2.71$, $p = .081$, $\eta_p^2 = 0.03$, indicating that the PJT decreased over time regardless of the order in which the responses and their corresponding behaviours were presented (separate analyses for each viewing direction condition indeed revealed that the effect of presented response was significant for both the direct, $F(1.58, 68.02) = 5.36$, $p = .011$, $\eta_p^2 = 0.11$, and the reverse condition, $F(1.66, 63.09) = 16.13$, $p < .001$, $\eta_p^2 = 0.30$). No other main effects or interactions were statistically significant. These findings support a cognitive rather than a behavioural influence on judgments.

The second ANOVA on accuracy scores revealed accuracy was greater for truths (*M* = .62, *SD* = .02) than for lies (*M* = .48, *SD* = .02), F (1, 81) = 20.64, $p < .001$, $\eta_p^2 = 0.20$. This was moderated by the presented response, F (1.62, 131.58) = 21.20, $p < .001$, $\eta_p^2 = 0.21$. When judging truths, accuracy was higher for the first than for the second, t (81) = 3.58, $p < .001$, $d = 0.34$, and third presented responses, t (81) = 4.05**,** $p < .001$, $d = 0.37$. In contrast, when judging lies accuracy was lower for the first than for the third presented response, t (81) = 3.79, $p < .001$, $d = -0.34$. No other main effects or interactions were significant.

Additional analyses. A decrease in bias could cause a decrease in accuracy. To separate the effects of accuracy and bias, we used signal detection measures. A 3 (Presented Response) x 2 (Viewing Direction) ANOVA on B''_D yielded a significant main effect of presented response, F (1.53, 123.82) = 17.96, $p < .001$, $\eta_p^2 = .181$. The truth bias declined between the first (*M* = .39, *SD* = .47) and both the second (*M* = .21, SD = .48), t (81) = 4.16, $p < .001$, $d = 0.38$, and third rating (*M* = .14, *SD* = .52), t (81) = 4.85, $p < .001$, $d = 0.50$, but not between the second and third ratings, t (81) = 2.10, $p = .105$, $d = 0.14$. The Presented Response x Viewing Direction interaction predicted by the behavioural account was not significant, F (1.53, 123.82) = 2.31, $p = .116$, $\eta_p^2 = 0.03$. A similar ANOVA on A' revealed no significant effects on accuracy.

We wanted to make sure the lack of a Presented Response x Viewing Direction interaction on the PJT reflected the real absence of an effect rather than a lack of statistical power. We calculated a Bayes Factor using a Cauchy prior distribution with a scaling factor for the fixed effects of 0.5 over the standardised effect sizes, and a scaling factor of 1.0 for the nuisance variables. We compared a complex model with the Presented Response x Viewing Direction interaction with a simpler model without this interaction (see the

Appendix for motivation of the scaling factor and for details on model specification). The analysis revealed that in order to prefer the more complex model we would need prior odds favouring it greater than about 45. This strongly supported the true lack of an interaction effect. A Bayes factor calculated in the same way for accuracy scores revealed that the data were 100 times more likely under the null hypothesis.

Summary. The bias decreased over successive judgments regardless of whether the statements were presented in the recorded order or in the opposite order, suggesting the decrease cannot be explained by changes in the speakers' behaviours over consecutive answers.

Testing the Default-Interventionist and Parallel-Competition Models

Because the data were not easily amenable to traditional *F*-tests, a model comparison approach was used to assess the effect of *cumulative viewing time* on bias and accuracy. Two generalised logistic mixed-effects models (GLMEMs) were created, one with all the manipulated variables and the other additionally including the fixed effect of Cumulative Viewing Time. A significant difference in the predictive ability of these two models would indicate that the addition of cumulative viewing duration significantly improved the fit of the data. It did not, neither for the PJT, χ^2 (1) = 0.03, p = .861, nor for accuracy, χ^2 (1) = 1.22, p = .290. The simpler model without viewing time is preferred.

Similarly, the duration of the *first presented response* could predict neither the PJT, χ^2 (1) = 0.77, p = .381, nor accuracy, χ^2 (1) = 0.67, p = .411, in judging that response.

Discussion

Experiment 1 established the change in bias is attributable to the rater, not to a change in the speakers' behaviour. Consistent with both HAM and step-by-step response mode explanations, there was initially a high truth bias that decreased over successive judgments, regardless of the order the responses were presented. Overall accuracy did not change over ratings.

Is the decrease in bias caused by the act of making multiple judgments (step-by-step account), or by the amount of time raters have to process the information (HAM account)? Consistent with research in persuasion (Thompson et al., 2003; see also Johnson-Laird & Byrne, 1993) and contrary to a time-based HAM, processing time could not predict how likely people were to believe speakers were telling the truth; thus, it appears that the act of rating over several occasions reduces the truth bias. However, the present findings do not allow us to dismiss HAMs altogether. We must consider a third class of HAMs identified by Evans (2007), *pre-emptive conflict resolution models*.

EXPERIMENT 2

Pre-emptive conflict resolution models do not propose that analytical processing will be seen only late in the judgment process. Instead, they propose a "decision" is made at the outset as to whether heuristic or systematic processing will be used (Evans, 2007). Different communication channels make different demands on cognitive resources, thereby making heuristic or analytic processing more likely. Visual cues are easier to process and require fewer cognitive resources; therefore, they can be processed heuristically (Reinhard, 2010; Reinhard & Sporer, 2008, 2010; Stiff et al., 1989). Verbal cues require greater cognitive resources (Gilbert & Krull, 1988); therefore, analytical processing is needed to process these cues (Chaiken, 1980; Reinhard, 2010; Reinhard & Sporer, 2010). If the truth bias results from heuristic processing, then visual cues should yield more of a *truth bias* than verbal cues (Burgoon, Blair & Strom, 2008). Further, because analytical processing takes a systematic approach towards forming judgments, *accuracy* should be higher when verbal cues (processed analytically) are available (Reinhard, 2010; Reinhard & Sporer, 2008).

We tested whether the findings from Experiment 1 would change depending on the communication channel. Data from the direct-viewing, audio-visual condition of Experiment 1 were compared with data from similar participants with access to only visual (video condition) or only audio (audio condition) information from the same videotape. Consistent with HAMs, we predicted that:

(a) More truth judgments would be made in the video (because heuristic processing would be engaged) than in the audio condition (systematic procession), with the audio-visual condition located between these.

(b) Accuracy would be lowest in the video condition and highest in the audio condition.

(c) The decrease in truth judgments would be weakest in the video condition – because switching to systematic processing would be difficult with no revealing verbal information available– and strongest in the audio condition.

(d) Accuracy would increase over consecutive judgments primarily in the audio condition, but not in the video condition.

Method

Psychology undergraduates were allocated to the video (n = 22; 15 female; age M = 20.55, SD = 4.18), audio (n = 27; 17 female; age M = 20.33, SD = 2.24) or audio-visual conditions (n = 24; 15 female; age M = 20.21, SD = 2.32; data of these participants came from Experiment 1; they were in the first group in the direct-viewing condition). Sex and age distributions did not differ between the groups.

The procedure closely followed Experiment 1, except for the modality manipulation and the fact that only the direct viewing direction videos were used.

Results

Truth Bias

A 2 (Veracity: truthful/deceptive statement) x 3 (Response: 1st/2nd/3rd) x 3 (Channel: video/audio/audio-visual) mixed ANOVA with repeated measures on the first two variables was run on the PJT. The PJT decreased over successive ratings F (1.74, 121.55) = 7.10, p = .002, η_p^2 = .092 (Table 1), but contrary to predictions, neither the channel main effect, F (2, 70) = 2.70, p = .074, η_p^2 = .072, nor the Response x Channel interaction were significant either, F (3.47, 121.55) = 1.10, p = .357, η_p^2 = .030. The reduction in the PJT was weakest in the video condition, as predicted, for which none of the pairwise comparisons were significant (see Table 1). A Response x Channel ANOVA run on B''_D confirmed these findings, and Bayes factors (with the data shifting plausibility towards the null by a factor of 18; see Appendix for details) provided strong evidence against the alternative hypotheses (these analyses are available from the first author).

Replicating Experiment 1, a GLMEM with maximal random effects determined that cumulative viewing duration could not add any predictive value to the model in fitting the PJT, χ^2 (1) = 0.07, p = .790, in either the video, χ^2 (1) = 0.17, p = .680, or audio channels, χ^2 (1) = 0.08, p = .772. Similarly, there was no significant effect of duration on the PJT to *the first* response, χ^2 (1) = 0.82, p = .364, regardless of whether only video, χ^2 (1) = 1.11, p = .293, or only audio information was present, χ^2 (1) = 0.82, p = .366. In each case, the simpler model should be preferred.

Accuracy

A similar ANOVA on accuracy revealed that raters were more accurate in judging truths (*M* = .58, *SD* = .02) than lies (*M* = .46, *SD* = .02), F (1, 70) = 14.90, $p < .001$, η_p^2 = .175. This was moderated by Response, F (1.73, 121.43) = 8.81, p = .001, η_p^2 = .112. Accuracy for lies increased between the 1st and the 2nd response but not further, while accuracy for truths decreased non-significantly throughout successive ratings (Table 2). Neither the channel main effect, F (2, 70) = 0.85, p = .431, η_p^2 = .024, nor the Channel x Response interaction, F (3.57, 125.06) = 0.78, p = .526, η_p^2 = .022, were significant. A Response x Channel ANOVA run on A' yielded no significant effects, and Bayes factors (shifting the odds in favour of the null by a factor of 2.59) provided evidence against the alternative hypotheses (analyses available from the first author).

GLMEM comparisons found that cumulative viewing duration could not predict accuracy, χ^2 (1) < 0.02, p = .878, in either the audio, χ^2 (1) < 0.01, p > .999, or video conditions, χ^2 (1) = 0.04, p = .840. Again, the duration of the first portion of the statement could not predict accuracy when rating the first response, χ^2 (1) = 2.12, p = .145. If only audio information was available, the duration of the first response could predict accuracy, χ^2 (1) = 5.90, B = 0.01, p = .015, but if only video information was presented it could not, χ^2 (1) = 1.21, p = .272.

Discussion

There was little support for a heuristic processing account of the truth bias: accuracy and bias were similar across all three cue-type conditions, and changes over successive ratings did not match HAM-based predictions (Evans, 2007). Also, the truth bias did not decline with longer viewing times, as would be expected by the two classes of HAM that claim a shift from heuristic to analytical processing over time. Equally, accuracy could not be predicted from viewing time, with the only exception of the audio condition. This latter finding may look consistent with an HAM, but HAMs explicitly predict a reduction in bias when switching to analytical processing; this was not supported across Experiments 1 and 2.

The present findings question both the behavioural account (Experiment 1) and the HAM account (Experiments 1 and 2), and are consistent with a step-by-step account: it is the number of judgments made, not the amount of processing time, what explains the shift in bias. In Experiment 3, we directly contrasted the predictions of the HAM and the step-by-step accounts.

EXPERIMENT 3

According to a HAM, when using very short clips bias would be high and accuracy low because time is needed for an analytical process to run to completion. For instance, interrupting participants' processing leads to truth biased responding (Gilbert et al., 1990), potentially indicative of heuristic processing. But according to a step-by-step account there should be a progressive decrease in bias regardless of whether the clips are long or short.

In Experiment 3, we used short (8 s) segments of each of the senders' responses. If the decrease in bias is detected with 8 s responses, this will be the result of making repeated judgments rather than of switching from heuristic to systematic processing. We also used a long clips control condition.

Method

Eighty-two undergraduates (63 female; age M = 19.29, SD = 3.22, range 18 - 36 years) participated in this experiment in the context of a Social Psychology lecture.

Materials and Procedure

The booklet and stimulus material used in Experiments 1 and 2 were employed. The procedure followed Experiment 1, except here all participants watched the speakers' responses in the recorded order and viewed either the full-length responses of each speaker (*long clips* condition, n = 49 participants) or only the first 8 s of each response (*short clips* condition, n = 33). Sex and age distributions did not differ significantly between the two conditions.

Results

Analyses were run on the PJT and B''_D to assess bias, as well as on accuracy scores and A' to assess accuracy. In order to avoid unnecessary reiteration and complexity, only the analyses of the signal detection theory measures are presented here. They were preferred because of the independence of B''_D and A' and because they are simpler. Both sets of analyses tell the same story. Missing analyses are available from the first author.

Truth Bias

A 3 (Response: 1st/2nd/3rd) x 2 (Clip Length: short/long) ANOVA on B''_D scores revealed a significant main effect of response, F (1.62, 129.58) = 4.63, p = .017, η_p^2 = 0.06. The truth bias declined form the first (M = .22, SD = .52) to the third response (M = .06, SD = .52), t (81) = 2.51, p = .042, d = 0.81. Neither the main effect of clip length nor the predicted Response x Clip Length interaction were significant, both ps > .143. Thus, response bias declined across the statement regardless of clip length.

A Bayes factor with all the variables in the preceding ANOVA revealed that in order to prefer a model with the interaction term over a model without it we would need prior odds greater than 2.9 favouring it. This offered moderate support for the null hypothesis of no interaction effect, supporting the step-by-step account.

Accuracy

Another Response x Clip Length ANOVA was conducted on A' scores. The response main effect was significant, F (1.97, 157.45) = 10.96, p < .001, η_p^2 = 0.12, but it also interacted with clip length, F (1.97, 157.45) = 9.81, $p < .001$, $\eta_p^2 = 0.11$. When rating long clips, accuracy did not change significantly across responses ($M = .60$, $SD = .20$; $M = .57$, $SD = .19$; and $M = .62$, $SD = .18$; for the 1st, 2nd, and 3rd response, respectively), all ps > .181. For short clips, there was an increase in accuracy from the first ($M = .45$, $SD = .20$) to the second response ($M = .59$, $SD = .19$), t (32) = 4.06, $p < .001$, $d = 0.72$, with no further increase (for the third response, $M = .63$, $SD = .18$). The main effect of clip length was not statistically significant, F (1, 80) = 1.12, $p = .294$, $d = 0.24$ (for short clips: $M = .56$, $SD = .16$; for long clips: $M = .60$, $SD = .15$).

In summary, response bias decreased across rating points, and clip length did not moderate this effect. Clip length influenced accuracy rates: with 8 s clips mean accuracy was below .50; longer viewing durations increased accuracy to approximately .60.

Discussion

The HAM claims analytical processing is slow and requires more time than heuristic processing (Evans, 2007). However, we found bias decreased over ratings, regardless of whether raters watched short or long clips. These findings support a step-by-step process.

Interestingly, accuracy increased for short clips (in particular from the first to the second response) but not for long clips. There are at least two explanations for the increased accuracy and reduced bias in the short clip condition. The raters may have switched from heuristic to systematic processing at some point between 8 s and 16 or 24 s. Alternatively, the low accuracy rate at the first rating point for the short clips condition may simply reflect that there was not enough information available at this point to make a reasoned judgment. Additional information provided during the second 8 s response may have permitted an increase in accuracy without necessarily reflecting a shift in processing modes. The thin slices were used to prevent switching to analytic processing claimed by Masip et al. (2009)

to require statements longer than 30 s, but we cannot rule out the possibility of analytical processing. However, because bias, our primary prediction arising from the HAM, could not be predicted by clip length, a non-HAM based account is preferred.

In any case, it is clear that a HAM account cannot explain the decrease in bias for long responses. If there is any switch between heuristic and systematic processing, it happened at some point between 8 and 32 s. The mean duration of the entire first response was 50 s, meaning analytical processing should already have been engaged and so a decrease in bias should not be seen. Therefore, as in Experiments 1 and 2, the HAM account cannot explain the decrease in bias. Instead, the data support a step-by-step process account. In Experiment 4, we considered a mechanism that could explain why step-by-step responding has this effect.

EXPERIMENT 4

Masip et al.'s (2009) study and Experiments 1 through 3 here evidenced a decline in truth bias between the first and second response, but no further decline. Why would step-by-step responding lead to this effect?

When a speaker makes multiple responses, it is possible to compare them. Raters use consistency more often than any other cue when comparisons can be made (Granhag & Strömwall, 1999, 2000a, 2001b). Consistency seemed a plausible candidate for explaining the decline in truth bias over time. Because raters perceive inconsistencies even when they are not present (Granhag & Strömwall, 2001b), raters could shift their judgments towards deception. Having established inconsistency between the first and second response, there may be no additional effect of continued perceived inconsistency by the third response. Thus, we predicted a greater decline in the PJT between the first and second response for speakers perceived as inconsistent than for those perceived as consistent.

Method

Participants

Consistency raters. Forty-nine undergraduates (40 females; age *M* = 19.04, *SD* = 2.78, range 17 - 30 years) rated the stimulus videos for verbal and nonverbal consistency.

Veracity judges. We examined the influence of perceived consistency on the lie-truth judgments of participants rating the full version of Masip et al.'s (2006) Video Set A1 in previous studies. This involved 14 raters in Masip et al.'s (2009) experiment, 83 raters of Experiment 1 above, and 49 raters in the long clip condition of Experiment 3, amounting to 146 undergraduates (117 female; age *M* = 20.35, *SD* = 2.76, range 18 - 36 years). Additionally, in Experiment 2, 22 undergraduates (15 female; age *M* = 20.55, *SD* = 4.18, range 18 to 38 years) were exposed to only visible information from the same video set, and 27 undergraduates (17 female; age *M* = 20.33, *SD* = 2.24, range 18 -28 years) were provided with only audio information.

Procedure

Consistency raters received an instruction sheet with definitions of *verbal consistency*, "the extent to which the same details or similar details are repeated over the responses with no contradiction", and *nonverbal consistency*, "the extent to which the same behaviours or similar behaviours are repeated over the responses". They watched Video Set A1 and provided ratings of verbal and nonverbal consistency for each speaker after viewing two responses, and then again after viewing all three responses. Ratings were given on a 1 (*Not consistent at all (Inconsistent)*) to 7 (*Fully consistent*) scale.

Consistency. The consistency ratings were used to median split the clips as high or low in consistency within each of the 2 (channel: verbal/nonverbal) x 2 (rating point: after the second (t_2) or third (t_3) response) cells. Ratings across the verbal and nonverbal channels were highly correlated. Cronbach's alphas were calculated separately for each of the Consistency x Rating Point x Veracity cells, and ranged between .90 and 1.00. Therefore, ratings were collapsed across channels.

Coding. Truth judgments were coded as 1 and lie judgments as 0. Then, the change in the PJT was calculated as the judgment at the second (or third) response minus the judgment at the first response. A shift from a truth (1) to a lie (0) judgment was coded as 0 – 1 = -1; a shift from a lie to a truth judgment was coded as 1 – 0 = 1, and no change in judgment was coded either as 0 – 0 = 0 or 1 – 1 = 0. Thus, a negative value indicates a *shift towards a lie* response and a positive value indicates a *shift towards a truth* response. This new variable, judgment change, was the dependent variable.

Results and Discussion

Diagnosticity

A 2 (Veracity: truthful/deceptive statement) x 2 (Rating Time: t_2/t_3) within subjects ANOVA revealed that truths were rated as more consistent (M = 4.50, SD = 0.74) than lies (M = 4.71, SD = 0.70), F (1, 48) = 5.24, p = .026, η_p^2 = 0.10.

Consistency use

Three 2 (Consistency: low/high) x 2 (Veracity: truthful/deceptive statement) x 2 (Rating Time: t_2/t_3) within-participants ANOVAs were conducted on the PJT change.[1] The first ANOVA was conducted on Masip et al.'s (2009) data, the second on the data of Experiment 1, and the third on the ratings of Experiment 3's full-length clips. The main effect of consistency was significant in all the three ANOVAs (Table 3; more detailed results are available from the first author). A meta-analysis of all three experiments yielded a weighted Hedges's unbiased g = -0.51, 95% CI [-0.73, -0.29], z = 4.46, p < .001, which is a medium effect size (Cohen, 1988). A homogeneity analysis showed that the sample of effect sizes was homogeneous, and hence variability was caused by sampling error alone, $Q = 2.37$, $p = .31$. In short, across the three studies consistency had a substantial impact on the decrease in the PJT (see Figure 1).

Consistency ratings were collected separately for verbal and nonverbal behaviour. We examine whether *nonverbal* consistency predicted the decline in the PJT in the video condition of Experiment 2, and whether *verbal* consistency did the same in the audio condition of Experiment 2. A Nonverbal Consistency x Veracity x Rating Time ANOVA did not yield any significant effect for the video condition of Experiment 2; however, a similar ANOVA conducted for audio-only clips revealed a significant main effect of verbal consistency, $F(1, 26) = 6.45$, $p = .017$, $\eta_p^2 = 0.20$. The decrease in truth judgments was stronger in verbally inconsistent ($M = -.10$, $SD = .16$) than in verbally consistent ($M = -.01$, $SD = .14$) statements (Figure 1).

In summary, the analyses of several studies and experimental conditions provide compelling evidence that consistency guides veracity judgments across multiple responses of the same sender. Verbal consistency may contribute more to this effect than nonverbal consistency.

Importantly, we do not claim that a consistency step-by-step account is necessarily orthogonal to a HAM account, but rather that the HAM account failed to explain the truth bias in three experiments and that the step-by-step account received support in all reported experiments.

GENERAL DISCUSSION

In making lie-truth judgments, raters show a bias towards believing (Bond & DePaulo, 2006). One line of research suggests that this truth bias is produced by a system that initially is biased towards believing with the short 30 s clips typically shown to lie detection raters, but with longer statements may use a more effortful evaluation (Masip et al., 2006, 2009, 2010; see also Gilbert, 1991). We tested the claims of these heuristic-analytic models (HAMs) as presented by Evans (2007).

We first considered whether the shift in judgments is a result of a shift in behaviour over the course of lengthier statements rather than the product of the rater's processing. The behavioural account was not supported. We then considered a stringent test of the HAM: the amount of processing time should predict the degree of bias or accuracy. Across three experiments we found this was not the case. In addition, we found that even when there was no initial bias (video condition of Experiment 2), raters were still less inclined to believe the speaker by the second and third judgments. These findings question default-interventionist or parallel-competition HAM explanations. Pre-emptive conflict resolution models, which claim an early selection of processing routes, could not explain the decline in truth bias either. The types of information available in Experiment 2 (visual, audio, or audio-visual) did not result in the predicted choice of heuristic or analytical processing from the outset. Instead, in all of these experiments the decrease in bias was better predicted by the mere act of making multiple judgments, in line with a step-by-step response mode account. Other research has shown that even at much larger time scales, from 1 to 5 months later, there is a decline in truth bias with subsequent ratings of the speaker's statement (Anderson, DePaulo, & Ansfield, 2002).

In Experiment 4, we sought to understand why step-by-step responding decreases the truth bias, and found that people use perceived consistency: when the speaker's responses appear inconsistent, raters shift away from a truth-biased position. This finding is in line with previous research. Granhag and Strömwall (2000b) found that 60% of the participants used consistency when rating a speaker's veracity from different statements made by the same speaker. Other research also shows that both practitioners and laypersons believe inconsistency indicates deception (Strömwall, Granhag, & Hartwig, 2004).

Here we found that, across four experiments, perceived inconsistency explained the decrease in truth judgments over the course of the speaker's statement. It seems therefore well established that raters repeatedly judging veracity during a statement make within-statement comparisons. In addition, we discovered that perceived inconsistency decreases the initial tendency to make truth judgments.

The findings are consistent with truth-default theory (Levine, 2014), which proposes people default to a truth belief unless they perceive a 'trigger' that leads them to consider the possibility of deception. In the current study, the trigger would be the perceived inconsistency in the statement.

The findings are also consistent with the smart lie detector account (Street & Richardson, in press; Street, 2014) that argues people make use of generalised rules (perceived inconsistency) to make informed judgments in low-diagnostic environments. In line with this, we found raters made use of perceived inconsistency, a diagnostic cue in this study. In this sense, it might be considered a smart heuristic (Gigerenzer, Todd & The ABC Research Group, 1999).[2]

The current findings are not intended to be an all-or-nothing challenge against the HAM account, but rather a piece of the puzzle. One possible limiting factor of the current findings is that we chose to focus on Evans' (2007) taxonomy of HAMs that make temporal predictions. However, analytical processing is not only slower than heuristic processing, but also non-automatic, requiring motivation and cognitive capacity. Other research has examined the role of motivation or cognitive capacity on lie detection or deception judgments (e.g., Millar & Millar, 1997; Reinhard & Sporer, 2008, 2010; Stiff, Kim, & Ramesh, 1992).

Our findings have practical implications both for researchers and lay citizens. Researchers investigating deception detection should use longer clips, as clips shorter than 87 s (average accumulated time by the end of Response 2) might bias judgments towards truthfulness and might provide a distorted picture of the participants' response tendencies and detection accuracy. Lay people should avoid hasty judgments in assessing veracity because they produce (truth-)biased judgments.

Conclusions

We focused on the processes involved in the truth bias. Truth judgments are often high initially, and then decline progressively over successive ratings. We showed that a HAM-based account could not explain either the initial truth bias or its decline. Instead, the reduced bias was attributable to the act of making multiple judgments, and reflected the use of a simple rule: seemingly inconsistent statements were less likely to be judged as truthful by the point of the second judgment.

References

Anderson, D. E., DePaulo, B. M., & Ansfield, M. E. (2002). The development of deception detection skill: A longitudinal study of same-sex friends. *Personality and Social Psychology Bulletin, 28*, 536-545.

Bond, C. F., Jr., & DePaulo, B. M. (2006). Accuracy of deception judgments. *Personality and Social Psychology Review, 10*, 214-234.

Burgoon, J. K., Blair, J. P., & Strom, R. E. (2008). Cognitive biases and nonverbal cue availability in detecting deception. *Human Communication Research*, *34*, 572-599.

Chaiken, S. (1980). Heuristic versus systematic information processing and the use of source versus message cues in persuasion. *Journal of Personality and Social Psychology*, *39*, 752-766.

Chaiken, S., Liberman, A., & Eagly, A. H. (1989). Heuristic and systematic information processing within and beyond the persuasion context. In J. S. Uleman & J. A. Bargh (Eds.), *Unintended thought* (pp. 212-252). New York: Guilford Press.

Chaiken, S., & Trope, Y. (Eds.). (1999). *Dual-process theories in social psychology*. New York: Guilford Press.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Donaldson, W. (1992). Measuring recognition memory. *Journal of Experimental Psychology: General*, *121*, 275-277.

Evans, J. S. B. T. (2007). On the resolution of conflict in dual process theories of reasoning. *Thinking & Reasoning, 13*, 321-339.

Evans, J. S. B. T., & Stanovich, K. E. (2013). Theory and metatheory in the study of dual processing: Reply to comments. *Perspectives on Psychological Science*, *8*, 263-271.

Gigerenzer, G., Todd, P. M., & The ABC Research Group (1999). *Simple heuristics that make us smart*. New York: Oxford University Press.

Gilbert, D. T. (1991). How mental systems believe. *American Psychologist, 46*, 107-119.

Gilbert, D. T. (1999). What the mind's not. In S. Chaiken & Y. Trope (Eds.), *Dual-process theories in social psychology* (pp. 3-11). New York: Guilford Press.

Gilbert, D. T., & Krull, D. S. (1988). Seeing less and knowing more: The benefits of perceptual ignorance. *Journal of Personality and Social Psychology, 54*, 193-202.

Gilbert, D. T., Krull, D. S., & Malone, P. S. (1990). Unbelieving the unbelievable: Some problems in the rejection of false information. *Journal of Personality and Social Psychology, 59*, 601-613.

Gilbert, D. T., Tafarodi, R. W., & Malone, P. S. (1993). You can't not believe everything you read. *Journal of Personality and Social Psychology*, *65*, 221-233.

Granhag, P. A., & Strömwall, L. A. (1999). Repeated interrogations: Stretching the deception detection paradigm. *Expert Evidence, 7*, 163-174.

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

Granhag, P. A., & Strömwall, L. A. (2000b). "Let's go over this again…": Effects of repeated interrogations on deception detection performance. In A. Czerederecka & T. Jaskiewicz-Obydzinska & J. Wojcikiewicz (Eds.), *Forensic Psychology and Law: Traditional Questions and New Ideas* (pp. 191-196). Krakow, Poland: Institute of Forensic Research Publishers.

Granhag, P. A., & Strömwall, L. A. (2001a). Deception detection based on repeated interrogations. *Legal and Criminological Psychology, 6*, 85-101.

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

Johnson-Laird, P. N., & Byrne, R. (1993). *Human reasoning: The psychology of deduction*. Hove, UK: Erlbaum.

Kahneman, D., & Frederick, S. (2002). Representativeness revisited: Attribute substitution in intuitive judgement. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgement* (pp. 49-81). Camrbdige, UK: Cambridge University Press.

Kassin, S. M. (2005). On the psychology of confessions: Does innocence put innocents at risk? *American Psychologist, 60*, 215-228.

Levine, T. R. (2014). Truth-default theory (TDT): A theory of human deception and deception detection. *Journal of Language and Social Psychology*, *33*, 378-392.

Levine, T. R., Park, H. S., & McCornack, S. A. (1999). Accuracy in detecting truths and lies: Documenting the "veracity effect". *Communication Monographs, 66*, 125-144.

Masip, J., Garrido, E., & Herrero, C. (2006). Observers' decision moment in deception detection experiments: Its impact on judgment, accuracy, and confidence. *International Journal of Psychology, 41*, 304-319.

Masip, J., Garrido, E., & Herrero, C. (2009). Heuristic versus systematic processing of information in detecting deception: Questioning the truth bias. *Psychological Reports, 105*, 11-36.

Masip, J., Garrido, E., & Herrero, C. (2010). Regression toward the mean or heuristic processing in detecting deception?: Reply to Elaad (2010). *Psychological Reports, 107*, 587-592.

Masip, J., & Herrero, C. (2013). "What would you say if you were guilty?" Suspects' strategies during a hypothetical behavior analysis interview concerning a serious crime. *Applied Cognitive Psychology, 27*, 60-70.

Millar, M. G., & Millar, K. U. (1997). The effects of cognitive capacity and suspicion on truth bias. *Communication Research, 24*, 556-570.

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

Morey, R. D., & Rouder, J. (2013). *BayesFactor: An R package for Bayesian analysis in common research designs*. Retrieved from http://bayesfactorpcl.r-forge.r-project.org

R Development Core Team (2011). *R: A language and environment for statistical computing.* Vienna, Austria: The R Foundation for Statistical Computing.

Rae, G. (1976). Table of A'. *Perceptual and Motor Skills*, *42*, 98.

Reinhard, M. A. (2010). Need for cognition and the process of lie detection. *Journal of Experimental Social Psychology, 46*, 961-971.

Reinhard, M. A., & Sporer, S. L. (2008). Verbal and nonverbal behaviour as a basis for credibility attribution: The impact of task involvement and cognitive capacity. *Journal of Experimental Social Psychology, 44*, 477-488.

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

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers, 31,* 137-149.

Stiff, J.B., Kim, H. J., & Ramesh, C. (1992). Truth biases and aroused suspicion in relational deception. *Communication Research, 19*, 326-345.

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

Street, C. N. H. (2013). *Lie detection: Cognitive processes*. (Doctoral dissertation, University College London, London, UK). Retrieved from http://discovery.ucl.ac.uk/1414942/

Street, C. N. H. (2014). *Humans as smart lie detectors*. Manuscript submitted for publication.

Street, C. N. H., & Richardson, D. C. (in press). Descartes versus Spinoza: Truth, uncertainty, and bias. *Social Cognition*.

Strömwall, L. A., Granhag, P. A., & Hartwig, M. (2004). Practitioners' beliefs about deception. In P. A. Granhag & L. A. Strömwall (Eds.), *Deception detection in forensic contexts* (pp. 229-250). Cambridge, UK: Cambridge University Press.

Stupple, E. J. N., & Ball, L. J. (2008). Belief-logic conflict resolution in syllogistic reasoning: Inspection-time evidence for a parallel processs model. *Thinking and Reasoning*, *14*, 168-189.

Thompson, V. A., Striemer, C. L., Reikoff, R., Gunter, R. W., & Campbell, J. I. D. (2003). Syllogistic reasoning time: Disconfirmation disconfirmed. *Psychonomic Bulletin & Review*, *10*, 184-189.

Trippas, D., Verde, M. F., & Handley, S. J. (2014). Using forced choice to test belief bias in syllogistic reasoning. *Cognition*, *133*, 586-600.

Vrij, A. (2008). *Detecting lies and deceit: Pitfalls and opportunities* (2nd ed.). Chichester, UK: Wiley.

Footnote

[1] To preserve the continuous nature of the data we also conducted a set of mixed effects model comparisons. The findings mirror those of the ANOVAs. These analyses are available from the first author on request.

[2] It is important to note here that a heuristic process, a proposed mechanism that is fast and engages in relatively effortless thought, must be differentiated from a heuristic, a simplified, rule built up from an individual's prior history with the world (Street, 2013). The findings suggest the use of a simple heuristic (consistency), but do not support a heuristic processing model.

Table 1

Mean (Standard Deviations) PJT in each Channel in Experiment 2

	Response 1	Response 2	Response 3
Video	$.51_{a}$ (.14)	$.48_{a}$ (.14)	$.50_{a}$ (.16)
Audio	$.61_{a}$ (.23)	$.55_{b}$ (.22)	$.56_{b}$ (.26)
Audiovisual	$.60_{a}$ (.24)	$.57_{ab}$ (.23)	$.54_{b}$ (.27)
Across Channels	$.58_{a}$ (.14)	$.53_{b}$ (.14)	53_{b} (.16)

Note. Means sharing a common subscript are not statistically different at $\alpha = .05$ according to Bonferroni-corrected post-hoc *t*-tests.

Table 2

Mean (Standard Deviations) Accuracy for Truths and Lies in Experiment 2

	Response 1	Response 2	Response 3
Lies	$.43_a$ (.21)	$.48_b$ (.17)	$.48_b$ (.20)
Truths	$.61_a$ (.02)	$.57_a$ (.02)	$.56_a$ (.21)

Note. Means sharing a common subscript are not statistically different at $\alpha = .05$ according to Bonferroni-corrected post-hoc *t*-tests.

Table 3

Mean Change in the PJT for the Low versus High Consistency Items, and Main Effects of Consistency in the Consistency x Veracity x Rating Time ANOVAs

	Low Consistency		High Consistency					
	Mean	*SD*	Mean	*SD*	*F*	df	*p*	η^2
Masip et al. (2009)	-.19	.15	-.01	.15	7.35	1, 13	.018	.36
Experiment 1	-.10	.17	-.03	.14	9.12	1, 82	.003	.10
Experiment 3	-.12	.18	-.03	.15	8.39	1, 48	.006	.15

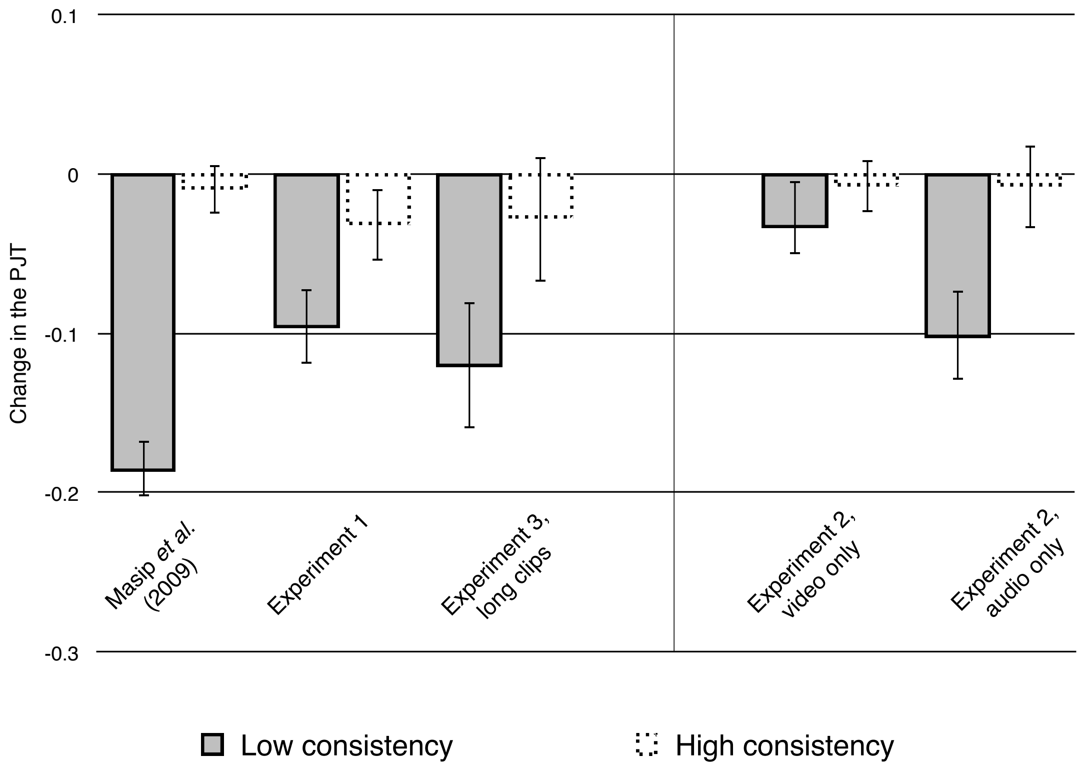
0.1
0
-0.1
-0.2
-0.3
Change in the PJT
Masip *et al.* (2009)
Experiment 1
Experiment 3, long clips
Experiment 2, video only
Experiment 2, audio only
Low consistency
High consistency

Figure Captions

Figure 1. The change in the PJT (overall) for high versus low consistency items, separated by experiment. Negative values indicate a shift away from a truth response. Whiskers denote one standard error.

Appendix

Some readers may be unfamiliar with some of the analytical approaches used in this article. These approaches are described below.

Signal Detection Theory (SDT) Measures

SDT measures calculate response bias and accuracy independently (Stanislaw & Todorov, 1999). We measured bias as B''_D (Donaldson, 1992) and accuracy as A' (Rae, 1976). B''_D ranges from -1 to +1, with 0 indicating no bias. Negative values reflect a bias toward responding lie and positive values a bias toward truth. A' is bounded between 0 and 1, with 0.5 reflecting chance accuracy.

Bayes Factors

The lack of significance in a statistical test could result from an underpowered study or from the lack of a real effect. The Bayes factor circumvents this issue by asking how probable one model versus another is, given the available data. Data can show support for or against the null or instead show the lack of an effect due to no evidence in either direction. Values near 1.0 indicate lack of power, while values of approximately 3 (and larger) indicate moderate evidence in favour of a specified (alternative or null) hypothesis. To illustrate, in Experiment 1 we compared two models: (a) a complex model with Veracity, Presented Response, Viewing Direction, and their interactions as fixed effects, with the PJT as the outcome variable, and with fully specified random effects for raters and speakers, and (b) a simpler model with the Presented Response x Viewing Direction interaction removed. In order to prefer the more complex model, we would need prior odds favouring it greater than about 45.

We used the BayesFactor package version 0.9.4 (Morey & Rouder, 2013) designed for the R statistical environment (R Development Core Team, 2011). A prior Cauchy scale of r = 0.5 over the effect sizes was selected; this prior includes 50% of the prior mass within the range of effect sizes between -0.5 and 0.5. This scaling factor is recommended by the BayesFactor documentation for most experimental designs given that it is readily computed and gives a stable integration of the likelihood. This relatively narrow prior is appropriate for the somewhat small effect sizes observed in lie detection research, and is used for all reported calculations. Fully specified random effects were included for raters and speakers in all analyses, as in the case of the generalised logistic mixed effects models described below.

Generalised Logistic Mixed Effects Models

Generalised logistic mixed effects models were used because the data were not always amenable to *F*-tests. We use the analysis from Experiment 1 on the cumulative viewing duration as an example here.

The fixed effect of Cumulative Viewing Time was added to a simpler model that included Veracity, Presented Response, and Viewing Direction, with the video-recorded speaker and the observer as random effects, each with its own random intercept. The random slopes for the speaker were Cumulative Viewing Time, Presented Response, and Viewing Direction. The random slopes for the observer were the Cumulative Viewing Time, Veracity and Presented Response. That is, slopes for all variables were permitted provided a slope was possible to model (i.e., provided the speaker or observer could be found in more than one cell for the given variable), resulting in a maximally specified mixed effects model. Restricted maximum likelihood estimates of the models were based on the Laplace approximation.