

## Lies, Damn Lies, and Expectations: How Base Rates Inform Lie–Truth Judgments

CHRIS N. H. STREET<sup>1\*</sup> and DANIEL C. RICHARDSON<sup>2</sup>

<sup>1</sup>Department of Psychology, University of British Columbia, Vancouver, Canada

<sup>2</sup>Cognitive, Perceptual and Brain sciences, University College London, London, UK

*Summary:* We are biased towards thinking that people are telling the truth. Our study represents the first test of how beliefs about the base rate of truths and lies affect this truth bias. Raters were told either 20, 50 or 80% of the speakers would be telling the truth. As the speaker delivered their statement, participants indicated moment by moment whether they thought the speaker was lying or being truthful. At the end of the statement, they made a final lie–truth judgment and indicated their confidence. While viewing the statement, base rate beliefs had an early influence, but as time progressed, all conditions showed a truth bias. In the final judgment at the end of the statement, raters were truth biased when expecting mostly truths but did not show a lie bias when expecting mostly lies. We conclude base rate beliefs have an early influence, but over time, a truth bias dominates. Copyright © 2014 John Wiley & Sons, Ltd.

### INTRODUCTION

People tend to believe others' statements to be truthful more often than they actually are, a phenomenon known as the truth bias (Bond & DePaulo, 2006). Although there is a substantial body of evidence demonstrating its influence (Levine, Park, & McCornack, 1999; Vrij, 2008), the source of the bias is relatively unexplored (Reinhard & Sporer, 2010). Here, we consider one possible account of the truth bias: as a reflection of the perceived base rate of honesty (DePaulo & Rosenthal, 1979; O'Sullivan, Ekman, & Friesen, 1988).

The truth bias has sometimes been considered as an error or inaccuracy in the judgment process (refer to Vrij, 2008). For example, we may willingly sacrifice accuracy and overlook others' lies in order to maintain social cohesion (O'Sullivan, 2003; refer also to Buller & Burgoon, 1996). Or, the bias may come about because of the structure of the mind. Gilbert, Krull, and Malone (1990) proposed that in order to comprehend, we first have no choice but to believe that the information is true; it is only afterwards that the information can be evaluated and possibly rejected. Alternatively, they may simply expect others to be generally decent and pleasant rather than conniving and deceptive, which can leave us open to being deceived (refer to Buller & Burgoon, 1996).

However, the truth bias may actually serve to *improve* accuracy (Levine, 2014; Street, 2013; refer also to Street & Richardson, ). The average person lies on a daily basis, but the number of deceptive interactions is far outweighed by the number of honest interactions we experience (DePaulo, Kashy, Kirkendol, Wyer, & Epstein, 1996; refer also to Serota, Levine, & Boster, 2010, for evidence that almost half the lies are told by a small percentage of people, suggesting that the proportion of deceptive encounters most people experience is particularly small). This may be because honesty is required for successful communication (Grice, 1975), because lying creates unpleasant and less intimate conversa-

tions (DePaulo et al., 1996), or because listeners begin from the default belief that others are generally honest (Levine, 2014). Whatever the reason, this 'bias' can serve as a useful aid to forming accurate judgments: Anticipating that others will be honest is likely adaptive in a world where people tell the truth more often than they lie.

There is some evidence supporting this position. Manipulations intended to make raters suspicious of deception make more lie judgments than naïve unsuspecting raters (e.g. Masip, Alonso, Garrido, & Herrero, 2009). In addition, raters show a bias towards disbelieving others when those others are believed to be sales people (P. J. DePaulo & DePaulo, 1989), which may reflect their perceptions of how honest these individuals are. Moreover, police investigators who show a lie bias (Meissner & Kassin, 2002) expect their interviewees to lie to them (Moston, Stephenson, & Williamson, 1992).

This seems to suggest that expectations influence the likelihood of believing others. However, it is not clear why. It may be because raters adjust their general expectations about all speakers as being more likely to be deceptive. Or, it may be at a more individual level: When judging a particular speaker, they may hedge towards disbelieving people when they are unsure. In the latter case, beliefs about how often people in general lie do not change. Instead, raters may ignore the base rate and be more willing to interpret individual behaviours as indicators of deceit. The research on suspiciousness cannot distinguish between these two possibilities. Here, we separate them and ask whether the truth bias is caused by beliefs about how often people in general lie and tell the truth.

Is there any evidence that lie detectors make use of the more general base rates of honesty? Although early research seemed to suggest that raters neglect the base rate (Tversky & Kahneman, 1974), in recent years, research has shown that it is taken into account (Barbey & Sloman, 2007). That early research found a base rate neglect may be a result of how that information was framed: Reframing the information so that it is informative and relevant reduces the neglect (Schwarz, Strack, Hilton, & Naderer, 1991). That is, raters ignore the base rate when the experiment is designed so that base rates are not informative or relevant.

\*Correspondence to: Chris N. H. Street, Department of Psychology University of British Columbia, Vancouver V6T 1Z4, Canada.  
E-mail: c.street@psych.ubc.ca

There is some evidence of sensitivity to base rate information from lie detection research. A study by Levine et al. (1999) manipulated the *actual* base rate of truths, although participants were not told anything about the likelihood of deception. They found that the truth bias remained constant regardless of the base rate. It is unclear whether this means that participants did not pick up on the true base rate, meaning it could not be used in their judgments or whether they did discover it, but that base rate information is not used when making lie–truth judgments. Similarly, in a set of experiments by Bond, Howard, Hutchison, and Masip (2013), when raters were only given information about speakers' incentives for lying or telling the truth (a perfectly diagnostic cue), raters' accuracy was near perfect. That is, base rates were perfectly defined by the incentives, and raters made use of that information when forming their judgments. However, when raters were able to observe the speakers' statements in addition to the incentive information, accuracy dropped markedly (Bond et al., 2013, Study 3). In fact, these judgments were not statistically more accurate than the highest accuracy rate obtained in research to that date where only the speakers' behaviours were available (Bond et al., 2013), potentially suggesting that raters ignored the base rate information when the speakers' behaviour was available.<sup>1</sup> It is not possible to distinguish between perceptions of the base rate and the actual base rate in this study: The incentive information, which may guide raters' *beliefs* about the base rate, was an accurate reflection of the *actual* base rate.

Here, we explicitly manipulated participants' beliefs about how often people lie and measured its effect on the truth bias. We predicted that if people expected mostly true statements, raters would show a greater bias towards judging statement as truthful. We also predicted that if people expected mostly deceptive statements, there would be a complimentary bias towards judging statements as deceptive. In addition, we predicted greater confidence in making 'truth' compared with 'lie' ratings when the majority of the statements were expected to be truthful, while the reverse was predicted when the majority of the statements were expected to be deceptive.

## METHODS

We presented rater participants with video clips of speakers lying and telling the truth about their holidays. Rater participants were told beforehand that either most speakers would lie, most would tell the truth, or that there would be an equal split of liars and truth tellers, although in reality, all participants saw an equal proportion of liars and truth tellers.

Participants were instructed to begin indicating whether they thought the statement was a lie or truth from video onset, and to continue responding all the time, the speaker was delivering their statement by holding down the response key or switching to the other key if they changed their mind.

<sup>1</sup> Elsewhere, we argue that the decline in accuracy shows that too much information can be a bad thing, for lie detectors. Specifically, attempts to form a single, categorical lie–truth judgment by combining multiple, probabilistic and noisy behaviours compound the potential for errors (Street & Richardson, 2014b).

They were told they could change their response as often or as little as they wished, to reflect what they felt at that moment about the statement. Finally, at the end of each video, participants provided a single lie–truth judgment of the statement they had just seen and their confidence in their judgment, as is typical of many lie detection experiments.

## MATERIALS

We used a stimulus set from a prior study (Street et al., 2011, April). Twenty-two participants were approached by a junior researcher posing as an assistant to a holiday guide documentary director. He determined a small set of countries the participants had and had not visited. The researcher said he was short of time and had been unable to find anyone who had visited some countries. So, as a favour, he asked participants to tell the director that they had spent time in one of the countries that they had never visited as well as give a monologue about a country they had truthfully visited. If participants agreed, they were taken to a live filming studio where they met and were left alone with another experimenter posing as the director. He stressed the importance of filming genuine accounts of their experiences in both countries. Participants signed a waiver to this extent stating that they would be entirely truthful in both their deliveries. The director was blind to which of the two statements would be deceptive, with the order of statements counterbalanced by the junior experimenter. Participants were asked to speak for approximately 30 s in response to the question, 'When you arrived in [country name], what was your first impression of the people there?'

Two speakers were excluded from the final stimulus set because they admitted to the director they were about to lie. That these participants were unable to maintain the ruse, we suggest, highlights our speakers truly believed they were about to mislead someone and that this would have consequences. Two further participants were selected as practice stimuli. This left 18 speakers each delivering a lie and truth, resulting in 36 videos. From these, two stimulus sets were created such that the 18 speakers appeared only once in each set and that each contained nine lies and nine truths. Truthful statements on average lasted 32.86 s ( $SD=10.79$ , median = 32.38 s, range 16 to 91 s). Deceptive statements on average lasted 32.72 s ( $SD=24.83$ , median = 23.87 s, range 10 to 55 s).

## Participants

Ninety-nine first year psychology students participated as part of their undergraduate studies. Fifteen participants were excluded because they had taken part in a study using the same stimulus set at an earlier date. One participant was excluded because of hearing difficulties, leaving 83 participants (67 females) with a mean age of 18.7 years ( $SD=0.70$  years, range 17 to 21 years).

## Procedure

Instructions explained that the speaker had provided a truthful and deceptive account about having visited a holiday

destination but that they would only see one of these accounts. Each video was counted in from three to one before onset. Throughout each statement, participants indicated moment by moment whether they currently believed or disbelieved the speaker by holding in either the left or right arrow keys while the speaker delivered their statement. Note that this was a categorical, dichotomous response of either lie or truth, not a scaled assessment of credibility. From the start of each statement, raters held down either a lie or truth response and could switch between the two responses as often or as little as they wished while the speaker was delivering the statement. Raters were told to indicate whether the entirety of the statement that they had seen up until that point made them believe the speaker was lying or telling the truth. In this way, a continuous measure of the lie–truth judgment trajectory was captured across the course of the statement.

After each video, participants indicated whether they believed that the speaker was lying or telling the truth by mouse clicking one of the options on screen. They also rated how confident they were in their response on a scale of 1 (not at all confident) to 10 (very confident).

Participants in the low truth expectancy condition ( $n=26$ ) were told that only ‘20% of the speakers were telling the truth that they had visited the country, 80% of speakers lied. That is, most people lied’. The high truth expectancy condition ( $n=27$ ) was told that 80% of speakers told the truth, and the equal split expectancy condition ( $n=30$ ) was told that half the speakers lied and half told the truth. The base rate information was presented both as a percentage with text explicitly explaining whether the majority of the speakers lied or told the truth. In reality, all conditions saw an equal split of lies and truths. Assignment of video set, position of lie–truth buttons on the screen, and question order (‘did the last person lie or tell the truth’ versus ‘did the last person tell the truth or lie?’) were counterbalanced.

## Design

Participants were randomly assigned to conditions. The independent variables were the anticipated proportion of truths to lies, i.e. either 20, 50 or 80% of truths were expected, and the veracity of the speaker’s statement. The dependent variables were the proportion of truth judgments (PTJ), confidence and accuracy taken after each statement had ended. This resulted in a 2 (veracity: lie or truth, within subjects)  $\times$  3 (expectation of truth: low, medium or high, between subjects) mixed design.

The continuous sampling of the judgment trajectory was also captured. The duration of a statement was split into five equal time points, resulting in a 2 (veracity)  $\times$  3 (expectation of truth)  $\times$  5 (time point:  $t_1$  to  $t_5$ , within subjects) mixed design. The PTJ and confidence responses were the dependent measures.

## RESULTS

While viewing the speakers, participants showed a lie bias when expecting mostly lies and a truth bias when expecting mostly truths. However, surprisingly, the effect of base rate beliefs at their final judgment, i.e. after the statement had finished, did not mirror the continuous rating of the statement

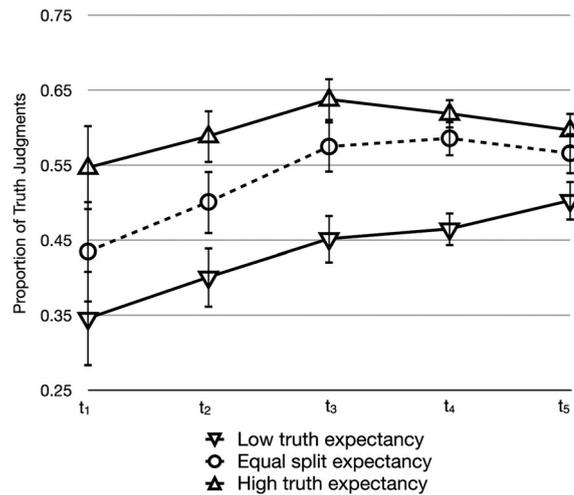


Figure 1. The proportion of truth judgments over the course of the average statement, split by beliefs about the base rate. Error bars denote one standard error

as it was being presented. Raters were most truth biased when expecting mostly truths. Interestingly though, expecting the majority of speakers to lie resulted in approximately half of all statements being judged as *truthful*.

Base rate information influenced confidence in much the same way. Participants were more confident in their truth than their lie judgments when expecting half or a majority of truths but were similarly confident in their lie judgments as their truth judgments when expecting mostly lies.

## Belief trajectory

Throughout the duration of each statement, raters indicated their belief at that moment in time. The continuous rating was binned into five proportional time points for ease of analysis. Within each bin, we calculated the proportion of time the truth response was given during that time period. A 2 (veracity, within subjects)  $\times$  3 (expectation of truth condition, between subjects)  $\times$  5 (time point, within subjects) ANOVA was conducted on the PTJ. A main effect of time was found,  $F(1.99, 113.52)=7.76$ ,  $p=.001$ ,  $\eta_p^2=0.12$ ,<sup>2</sup> showing significant linear,  $F(1, 57)=10.95$ ,  $p=.002$ ,  $\eta_p^2=0.16$ , and quadratic components,  $F(1, 57)=5.51$ ,  $p=.022$ ,  $\eta_p^2=0.09$ . There was also a main effect of veracity,  $F(1, 57)=4.32$ ,  $p=.042$ ,  $\eta_p^2=0.07$ , such that truths received a significantly higher PTJ ( $M=0.55$ ,  $SD=0.18$ , 95% CI [0.50, 0.60]) than lies ( $M=0.49$ ,  $SD=0.14$ , [0.46, 0.53]).

In line with our prediction, a main effect of condition was observed,  $F(2, 57)=9.64$ ,  $p<.001$ ,  $\eta_p^2=0.25$  (Figure 1). Planned post hoc  $t$  tests found the low truth expectancy condition ( $M=0.43$ ,  $SD=0.12$ , 95% CI [0.38, 0.49]) received a lower PTJ than the equal split ( $M=0.53$ ,  $SD=0.12$ , [0.47, 0.59]) and high truth expectancy conditions ( $M=0.60$ ,  $SD=0.12$ , [0.55, 0.65]),  $t(42)=-2.41$ ,  $p=.018$ ,  $d=-0.83$ , and  $t(35)=-4.34$ ,  $p<.001$ ,  $d=-1.42$  respectively. The equal split and high truth expectancy conditions showed no

<sup>2</sup> The assumption of sphericity was violated for the within-subject effect of time point, and so a Greenhouse-Geisser correction was used.

evidence of differing significantly,  $t(40) = -1.69$ ,  $p = .096$ ,  $d = -0.58$ . All interaction terms were non-significant,  $F_s < 1$ .

Finally, three one-sample  $t$  tests compared the degree of bias to the chance PTJ, 0.50. The low truth expectancy condition showed a lie bias,  $t(18) = 2.96$ ,  $p = .008$ ,  $d = 0.68$ . The equal split condition showed no evidence of a significant bias,  $t(16) = -1.24$ ,  $p = .234$ ,  $d = -0.30$ . Moreover, the high truth expectancy condition showed a truth bias,  $t(23) = -3.29$ ,  $p = .003$ ,  $d = -0.67$ . These findings are in line with our hypothesis.

### Truth bias as base rate belief

After rating throughout the trial, participants provided a single lie–truth judgment of the statement as a whole. A 2 (veracity, within subject)  $\times$  3 (expectation of truth condition, between subjects) mixed ANOVA was conducted on the PTJ. A main effect of veracity was found,  $F(1, 80) = 64.64$ ,  $p < .001$ ,  $\eta_p^2 = 0.45$ : Truthful statements received a higher PTJ than lies (Table 1). The predicted main effect of expectation was also present,  $F(2, 80) = 24.30$ ,  $p < .001$ ,  $\eta_p^2 = 0.38$ . Planned  $t$  tests found that the high truth expectancy condition received a higher PTJ than the equal split expectancy and low truth expectancy conditions,  $t(80) = 3.47$ ,  $p < .001$ ,  $d = 1.32$  and  $t(80) = 4.71$ ,  $p < .001$ ,  $d = 1.79$  respectively. The equal split and low truth expectancy conditions exhibited a medium effect size in the predicted direction, although there was no evidence of a statistically significant difference at the  $\alpha = .05$  level,  $t(80) = 1.26$ ,  $p = .079$ ,  $d = 0.49$ .

The veracity  $\times$  expectation interaction was also significant,  $F(2, 80) = 3.20$ ,  $p = .046$ ,  $\eta_p^2 = 0.07$ . Bonferroni-corrected post hoc  $t$  tests showed that when rating lies, the high truth expectancy condition produced a significantly higher PTJ than both the equal split,  $t(56) = 5.27$ ,  $p < .001$ ,  $d = 1.45$ , and low truth expectancy conditions,  $t(55) = 5.57$ ,  $p < .001$ ,  $d = 1.56$ , but the equal split and low truth expectancy conditions did not show evidence of differing from each other,  $t(52) = 0.34$ ,  $p > .999$ ,  $d = 0.01$ .

Truthful statements received a significantly different PTJ in each of the three conditions, with the high truth expectancy condition receiving a higher PTJ than both the equal split,  $t(56) = 2.64$ ,  $p = .031$ ,  $d = 0.72$ , and low truth expectancy conditions,  $t(55) = 5.26$ ,  $p < .001$ ,  $d = 1.55$ , and the equal split condition receiving a higher PTJ than the low truth expectancy condition,  $t(52) = 2.61$ ,  $p = .031$ ,  $d = 0.71$ . That is, ratings of truths followed the predicted decline in PTJ in line with the raters' expectations. However, when rating lies the decline in PTJ from high to low expectancy did not follow predictions, with participants in the low truth expectancy making a comparable degree of truth judgments as those in the equal split expectancy condition,  $t(52) = -0.30$ ,  $p > .999$ ,  $d = 0.08$ . Those in the high truth expectancy

condition made more truth judgments when rating deceptive videos compared with those equal split expectancy conditions when rating deceptive videos,  $t(56) = 5.27$ ,  $p < .001$ ,  $d = 1.40$  (Table 1).

### Confidence

Replicating much past research (Aamodt & Custer, 2006), confidence did not correlate with accuracy,  $r(81) = 0.06$ ,  $p = .615$ . We were concerned with whether anticipating a majority of truthful statements leads to higher confidence in making a truth judgment and whether anticipating a majority of lies leads to higher confidence in making lie judgments. A 2 (judgment veracity, within subjects)  $\times$  3 (expectancy condition, between subjects) mixed ANOVA was conducted on the confidence scores. There was no main effect of condition,  $F(2, 80) = 0.32$ ,  $p = .731$ ,  $\eta_p^2 = 0.01$ : There was no evidence that participants felt any more or less confident solely as a result of differing expectations. There was a main effect of veracity,  $F(1, 80) = 14.48$ ,  $p < .001$ ,  $\eta_p^2 = 0.15$ . Raters were more confident in making truth judgments ( $M = 4.59$ ,  $SD = 0.81$ , 95% CI [4.41, 4.76]) than lie judgments ( $M = 4.28$ ,  $SD = 0.85$ , [4.10, 4.46]), as found in prior research (DePaulo, Charlton, Cooper, Lindsay, and Muhlenbruck, 1997).

In line with our prediction, there was a veracity  $\times$  expectancy condition interaction,  $F(2, 80) = 16.02$ ,  $p < .001$ ,  $\eta_p^2 = 0.29$  (Figure 2). Planned  $t$  tests found no evidence that raters in the low truth expectancy condition were any more or less confident in making lie than truth judgments, with a small effect size observed,  $t(25) = 1.85$ ,  $p = .069$ ,  $d = 0.32$ . Both the equal split,  $t(26) = 2.39$ ,  $p = .019$ ,  $d = 0.41$ , and the high truth expectancy conditions,  $t(29) = 6.34$ ,  $p < .001$ ,  $d = 1.03$ , were more confident in making truth ratings than lie ratings.

We were also interested in the shift in confidence between conditions in addition to between lie–truth judgments. Bonferroni-corrected  $t$  tests indicated that when making lie judgments, there was a greater confidence in making lie judgments in the low truth expectancy condition compared with the high truth expectancy condition,  $t(55) = 2.48$ ,  $p = .045$ ,  $d = 0.67$ , all other  $t_s < 1.96$ . Similarly, when making truth judgments, there was greater confidence in the high compared with the low truth expectancy condition,  $t(55) = 2.56$ ,  $p = .037$ ,  $d = 0.68$ , but no other comparisons were significant,  $t_s < 1.96$ .

### DISCUSSION

People tell the truth far more often than they lie (DePaulo et al., 1996). In fact, half the lies that are told are delivered by a small minority of people (Serota et al., 2010). Perhaps, it is unsurprising, then, that raters tend to judge statements as truthful more often than as deceptive. This article reports the

Table 1. Mean proportion of truth judgments ((standard deviation), [95% confidence intervals]) at the end of the statement

	Low truth expectancy	Equal split expectancy	High truth expectancy	Total
Lie	.41 <sub>a</sub> (.17), [.35, .48]	.42 <sub>a</sub> (.18), [.36, .49]	.66 <sub>b</sub> (.15), [.60, .72]	.50 (.16), [.46, .53]
Truth	.55 <sub>a</sub> (.16), [.49, .61]	.67 <sub>b</sub> (.18), [.61, .73]	.79 <sub>c</sub> (.15), [.72, .84]	.67 (.16), [.63, .70]
Total	.48 (.13), [.43, .53]	.55 (.13), [.50, .60]	.72 (.13), [.67, .77]	

Means in each row sharing a common subscript are not statistically different at  $\alpha = .05$  after Bonferroni corrections.

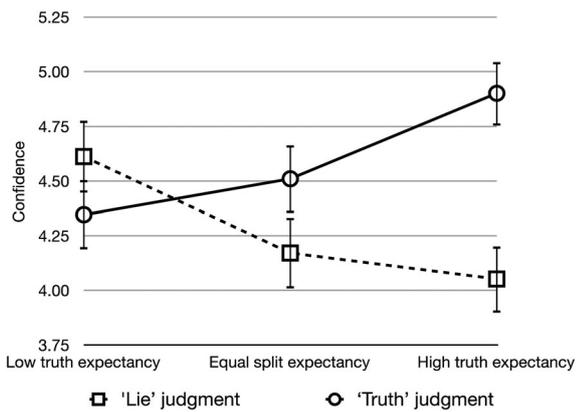


Figure 2. Confidence when making judgments of 'lie' or 'truth' as a function of base rate beliefs. Error bars denote one standard error

effects of the *perceived* base rate of honesty on the tendency to believe others. Our findings indicate that the truth bias can be moderated by beliefs about the base rate. Across the course of the statements and the raters' developing judgments, there was an effect of base rate beliefs in the predicted direction: Raters showed a lie bias when expecting mostly lies and a truth bias when expecting mostly truths. In particular, there was an early effect of base rate beliefs, but over time, raters shifted towards believing the speakers.

The lie–truth final judgments of the statements reflect this shift towards believing. There was a greater truth bias observed when raters expected mostly truths compared with when they expected an equal split of lies and truths. Raters in both conditions expressed more confidence in making truth compared with making lie judgments.

However, despite expecting every four out of five speakers to deceive them, raters in the low truth expectancy condition surprisingly showed a similar rate of truth responding to those in the equal split base rate belief condition. Raters in the low truth expectancy condition were also no more or less confident in their truth ratings than their lie ratings.

It might appear that the belief trajectory is fundamentally distinct from the final lie–truth judgment, given the differences in the degrees of bias between each. The difference can be accounted for by the fact that there was a shift towards believing speakers over time (regardless of manipulation). Had the time course of the developing judgment not been captured at all, it may have been tempting to conclude that believing that most speakers would lie has little effect on the degree of truth bias. Instead, our findings indicate an early effect of perceptions of the base rates. Our novel method of capturing the developing lie–truth judgment shows that there can be influences on the judgment that are not readily apparent by collecting a single lie–truth judgment at the end of the statement.

Base rate beliefs may have had an early effect because there was little information available from the speaker at the initial moments. As a result, raters may have used other available evidence, in this case their beliefs about the base rate. That is, in the absence of proximal information about the speaker's honesty (their behaviour and statement), raters may come to rely on context-relevant information (Street, 2013). We are currently following up this possibility. The

initial rating based on context (base rate beliefs) may have served as an anchor from which the belief developed (Fan, Wagner, & Manstead, 1995).

This finding is difficult to align with one the Spinozan mind account of the truth bias (Gilbert et al., 1990). This account claims 'all ideas are accepted... *prior* to rational analysis of their veracity, and that some are subsequently *unaccepted*' (Gilbert et al., 1990, italics in original). That is, raters initially and automatically assert belief but only afterwards can they evaluate the information and tag it as false if there is reason to doubt. The second stage requires both cognitive resources, because negation is thought to require greater effort, and more processing time, so that the process can develop to the second evaluative stage (Gilbert, 1991; Gilbert, Tatarodi, & Malone, 1993; refer also to Hasson, Simmons, & Todorov, 2005, and Masip, Garrido, & Herrero, 2006). Studies have tested the account by either making the task more cognitively taxing (e.g. by interruption: Gilbert et al., 1990; Hasson et al., 2005) or by manipulating the amount of processing time available (e.g. Gilbert et al., 1993; Skurnik, Yoon, Park, and Schwarz, 2005). With little processing time available, people show the Spinozan belief bias (Gilbert et al., 1993, Study 3), even if raters are forewarned (Gilbert et al., 1990, Study 3). In our study, we found that raters were able to take on a sceptic set from the earliest moments of consideration when they anticipated that most of the speakers would lie. Early in the judgment process, raters rely on their prior knowledge and understanding of the context to make the most informed guess (Street, 2013; Street & Richardson, in press), and sometimes, this can appear anti-Spinozan.

An interesting and unexpected aspect of our data is worth discussing. Regardless of base rate manipulations, all groups became more truth biased over time. Although we should be cautious about interpreting this finding until it is replicated with another stimulus set, there are at least four accounts of this increase that we tentatively suggest: (1) a behavioural self-presentational account, (2) a slow truth-biased mechanism, (3) a shift towards using their own base rate knowledge of the world outside the laboratory, and (4) an understanding of the social rules of politeness that discourages labelling others liars.<sup>3</sup>

First, given that liars want to be seen to be telling the truth, their self-presentational behaviour may have successfully persuaded raters of their honesty (refer to DePaulo et al., 2003). The shift towards a truth-biased response may be attributable to increasingly availability of apparently honest behaviour from both deceptive and honest speakers.

A second explanation is that the initial use of base rate information is followed by the onset of some distinct but unknown cognitive or emotional process. The unknown process is the root cause of the truth bias. This account suggests that the truth bias is a cognitive or emotional phenomenon on part of the rater rather than attributable to the self-presentational behaviour of the speaker and that it operates at a later timescale than base rate information. One possible such mechanism draws on the Truth Default Theory (TDT: Levine, 2014). TDT argues people work on the assumption that others are honest, but that suspicion

<sup>3</sup> We would like to thank an anonymous reviewer for this suggestion

can at least temporarily lead to an abandoning of the presumption of others' honesty, provided certain triggers are available. One such trigger may be information about the base rate.

This at first seems consistent with our findings of an initial bias towards disbelieving in the low truth base rate belief condition, which wanes over time towards a truth belief by the end of the statement. This could suggest a gradual shift back to the default state of believing others. However, the increase over time is only a slice of our experiment, showing the average trial. At the end of one trial, having shifted back to the default state, our participants would have to muster their temporary suspiciousness anew with each trial of sufficient magnitude to show active disbelief in the speaker. There is no trigger between trials reminding participants of the low base rates of honesty, so there seems to be no warrant to participants' abandoning their truth default state somewhere between each trial. Another way TDT could try to account for these findings is by assuming that raters maintain a high level of suspicion across the experiment but that in general speakers have an honest demeanour, which becomes more evident as the speaker continues to deliver their statement (although refer to Street & Masip, 2014, for evidence that the degree of truth bias over time appears to be more a function of the raters' cognitions than of the speakers' behaviours).<sup>4</sup>

A third account of the increasing truth bias makes a distinction between the 'sample' base rate and the 'population' base rate. By sample base rate, we mean the information offered to participants in this study regarding the small sample of speakers they were to rate. This varied between conditions. By population base rate, we mean the information participants bring to the task about the base rate of honesty from their entire history of interactions in their daily lives. The rate of honesty in real-world interactions is far greater than the rate of deception experienced (e.g. DePaulo et al., 1996; Serota et al., 2010; refer also to O'Sullivan et al., 1988). As a statement progresses, raters may be more influenced by their experiential knowledge of the population base rate and shift towards a truth belief.

Finally, a social level accusatory reluctance account could explain our findings (O'Sullivan, 2003). Throughout the statement, raters provided a continuous response, but they did not have to commit to that judgment: They were allowed to change their response as often as they wished. Perhaps, as the end of the statement approached, when raters were aware they would be required to commit to calling the speaker a liar or truth teller, they felt reluctant to break with the social rules of politeness and to commit to labelling the speaker a liar.

These results show the advantage of empirically investigating beliefs about the base rate effect, rather than simply assuming its presence. It is too early to attempt to extrapolate our findings to applied domains, but if our findings stand the test of time, legal professionals should bear in mind that their beliefs about the frequency of lying could influence their

judgments. A sceptic view may result in a tendency towards disbelieving when listening to a statement.

At the point of the final judgment, our student sample showed a truth bias regardless of their base rate beliefs. Law enforcement professionals have been shown to exhibit a lie bias (Meissner & Kassin, 2002). It is unclear whether our findings show that raters' pre-existing biases can be amplified (but not reversed) at the point of the final judgment or whether; regardless of raters pre-existing truth or lie biases, a truth bias can be formed based on base rate information but a lie bias cannot. An important question for future research, then, is if police investigators will show an increased lie bias if they believe that most of the speakers will lie and a dampened lie bias if they believe that most speakers will tell the truth.

In summary, we examined the effects of perceived base rates on the truth bias. Over the course of the statement, their developing judgment was in line with the perceived base rates, but by the point of their final judgment, there was no direct one-to-one relationship between the expected proportion of deceptive and honest statements and the actual proportion of lie and truth judgments. We observed an increasing shift towards believing the statement to be true over time prior to the final judgment, although it is unclear why this shift occurred. Four potential explanations were offered to encourage further exploration, ranging from behavioural to social level accounts. Our study shows that base rate expectations (independent of the actual base rate) do influence the trajectory of the belief and the final lie-truth judgment, resulting in an early lie bias when expecting mostly lies and an early truth bias when expecting mostly truths. Showing a bias towards the perceived base rate is better seen as a normative use of information rather than a biased and inaccurate view of the world.

## REFERENCES

- Aamodt, M. G., & Custer, H. (2006). Who can best catch a liar? A meta-analysis of individual differences in detecting deception. *The Forensic Examiner*, 15, 6–11.
- Barbey, A. K., & Sloman, S. A. (2007). Base-rate respect: From statistical formats to cognitive structures. *Behavioral and Brain Sciences*, 30, (3), 287–297.
- Bond, C. F., Howard, A. R., Hutchison, J. L., & Masip, J. (2013). Overlooking the obvious: Incentives to lie. *Basic and Applied Social Psychology*, 35, 212–221.
- Bond, C. F., & DePaulo, B. M. (2006). Accuracy of deception judgments. *Pers. Soc. Psychol. Rev.*, 10, (3): 214–234.
- Buller, D. B., & Burgoon, J. K. (1996). Interpersonal deception theory. *Commun. Theory*, 6, (3), 203–42.
- DePaulo, B. M., Charlton, K., Cooper, H., Lindsay, J. J., & Muhlenbruck, L. (1997). The accuracy-confidence correlation in the detection of deception. *Pers. Soc. Psychol. Rev.*, 1, (4): 346–357.
- DePaulo, B. M., Kashy, D. A., Kirkendol, S. E., Wyer, M. M., & Epstein, J. A. (1996). Lying in everyday life. *J. Pers. Soc. Psychol.*, 70, (5), 979–995.
- DePaulo, B. M., Lindsay, J. J., Malone, B. E., Muhlenbruck, L., Charlton, K., & Cooper, H. (2003). Cues to deception. *Psychol. Bull.*, 129, (1) 74–118.
- DePaulo, B. M., & Rosenthal, R. (1979). Telling lies. *J. Pers. Soc. Psychol.*, 37, (10), 1713–1722.
- DePaulo, P. J., & DePaulo, B. M. (1989). Can deception by salespersons and customers be detected through nonverbal behavioral cues? *J. Appl. Soc. Psychol.*, 19(18), 1552–1577.
- Fan, R. M. T., Wagner, H. L., & Manstead, A. S. R. (1995). Anchoring, familiarity, and confidence in the detection of deception. *Basic and Applied Social Psychology*, 17(1–2), 83–96.

<sup>4</sup> Importantly, we are not claiming that our data contradict TDT; merely that it is not clear to us, as the theory stands, how it would try to account for our data.

- Gilbert, D. T., Krull, D. S., & Malone, P. S. (1990). Unbelieving the unbelievable: Some problems in the rejection of false information. *J. Pers. Soc. Psychol.*, *59*(4), 601–613.
- Gilbert, D. T., Tafarodi, R. W., & Malone, P. S. (1993). You can't not believe everything you read. *J. Pers. Soc. Psychol.*, *65*(2), 221–233.
- Gilbert, D. T. (1991). How mental systems believe. *American Psychologist*, *46*(2), 107–119.
- Grice, H. P. (1975). Logic and conversation. In P. Cole & J. Morgan (Eds.), *Syntax and Semantics* (Vol. 3, pp. 41–58). New York: Academic Press.
- Hasson, U., Simmons, J. P., & Todorov, A. (2005). Believe it or not: On the possibility of suspending belief. *Psychol. Sci.*, *16*(7), 566–571.
- Levine, T. R. (2014). Truth-default theory (TDT): A theory of human deception and deception detection. *Journal of Language and Social Psychology*, *33*(3), 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*(2), 125–144.
- Masip, J., Garrido, E., & Herrero, C. (2006). Observers' decision moment in deception detection experiments: Its impact on judgment, accuracy, and confidence. *Int. J. Psychol.*, *41*(4), 304–319.
- Masip, J., Alonso, H., Garrido, E., & Herrero, C. (2009). Training to detect what? The biasing effects of training on veracity judgments. *Applied Cognitive Psychology*, *23*(9), 1282–1296.
- Meissner, C. A., & Kassin, S. M. (2002). "He's guilty!": Investigator bias in judgments of truth and deception. *Law Hum. Behav.*, *26*(5), 469–480.
- Moston, S., Stephenson, G. M., & Williamson, T. M. (1992). The Effects of Case Characteristics on Suspect Behavior during Police Questioning. *British Journal of Criminology*, *32*(1), 23–40.
- O'Sullivan, M. (2003). The fundamental attribution error in detecting deception: The boy-who-cried-wolf effect. *Pers. Soc. Psychol. Bull.*, *29*(10), 1316–1327.
- O'Sullivan, M., Ekman, P., & Friesen, W. V. (1988). The effect of comparisons on detecting deceit. *Journal of Nonverbal Behavior*, *12*(3), 203–215.
- Reinhard, M. A., & Sporer, S. L. (2010). Content versus source cue information as a basis for credibility judgments: The impact of task involvement. *Soc. Psychol.*, *41*(2), 93–104.
- Schwarz, N., Strack, F., Hilton, D. J., & Naderer, G. (1991). Base rates, representativeness, and the logic of conversation: The contextual relevance of "irrelevant" information. *Social Cognition*, *9*, 67–84.
- Serota, K. B., Levine, T. R., & Boster, F. J. (2010). The prevalence of lying in America: Three studies of self-reported lies. *Human Communication Research*, *36*, 2–25.
- Skurmik, I., Yoon, C., Park, D. C., & Schwarz, N. (2005). How warnings about false claims become recommendations. *Journal of Consumer Research*, *31*(4), 713–724.
- Street, C. N. H. (2013). Lie detection: Cognitive processes. (Doctoral dissertation). UCL (University College London), London, UK. Retrieved from Discovery UCL [accessed number 1414942].
- Street, C. N. H., & Masip, J. (2014). The source of the truth bias: A heuristic process? Manuscript submitted for publication.
- Street, C. N. H., & Richardson, D. C. (in press). Descartes versus Spinoza: Truth, uncertainty and bias. *Social Cognition*.
- Street, C. N. H., & Richardson, D. C. (2014b). The focal account: Does indirect lie detection research really access unconscious, implicit knowledge? Manuscript submitted for publication.
- Street, C. N. H., Tbaily, L., Baron, S., Khalil-Marzouk, P., Hanby, B., Wright, K., & Richardson, D. C. (2011). *Bloomsbury deception set*. Paper presented at the British Psychological Society Division of Forensic Psychology Conference.
- Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*(4157), 1124–1131.
- Vrij, A. (2008). *Detecting Lies and Deceit: Pitfalls and Opportunities* (2nd ed.). Chichester: John Wiley.