Detecting deception by loading working memory*

Hannah Faye Chua
University of Michigan

Richard E. Nisbett
University of Michigan

Jason Buhle
Columbia University

Katherine Rice
University of Michigan

Daniel Osherson
Princeton University

January 28, 2009

Abstract

Compared to truthful answers, deceptive responses to queries are expected to take longer to initiate. Yet attempts to detect lies through reaction time (RT) have met with limited success. We describe a new procedure that seems to increase the RT difference between truth-telling and lies. It relies on a Stroop-like procedure in which responses to the labels true and false are sometimes reversed. The utility of this method is assessed in a laboratory study involving both statements of fact and attitude. Three different ethnic groups participated: Americans of European ancestry, East Asians, and native Arabic speakers. For all three groups, our technique allowed identification of prevarication reliably better than chance.

*Contact: hchua@umich.edu.

Introduction

It seems plausible that fabricating a lie is more cognitively demanding than reporting the truth (Zuckerman et al., 1981). It may therefore be expected that deceptive communications will tend to exhibit response properties associated with high cognitive load, notably, elevated reaction time (RT) when subjects are instructed to respond as quickly as possible. Yet a recent meta-analysis of paraverbal indicators of deception (Sporer and Schwandt, 2006) reveals only a modest relation between veracity and RT.

Now suppose that response selection is rendered more difficult by the addition of a processing stage that is common to both lying and truth-telling. The common increment in processing may not affect cognitive load linearly. In particular, the impact on lying might be greater than for truth-telling if working memory resources are close to depletion for prevarication. In this case, behavioral measures like RT might show greater separation for deceptive compared to truthful responding.

We pursue this strategy in the experimental study described below. The processing stage added to all responses (both deceptive and truthful) is *truth-reversal*: on some trials, subjects must respond false to statements they mean to affirm (either truthfully or deceptively) and true to statements they mean to deny. Truth-reversal thus imposes a response conflict that is present in many forms of the Stroop phenomenon (Stroop, 1935; MacLeod, 1991). It is known that such incongruent responding activates the anterior cingulate cortex (Pardo et al., 1990). The latter structure appears to monitor performance and signal when adjustments in control are needed (MacDonald et al., 2000). Its response to Stroop-like conflict might especially interfere with the reversal needed to switch between affirmation and denial when prevaricating; the impact of truth-reversal would thus be greater under deception. Let us attempt to motivate such enhanced interference for lies.

We may speculate that truth-reversal affects prevarication more than truth-telling because it requires the same cognitive operation to be done then undone. Specifically, prevarication under truth-reversal demands a switch between affirmation and denial (the lie) followed by the opposite switch (truth-reversal). To illustrate, consider a Republican pretending to be a Democrat. Faced with "I am a Republican" under truth-reversal, prevarication leads to denial of the statement followed by reaffirmation due

to reversal. The special difficulty of revisiting an operation already executed would be analogous to *inhibition of return* (Klein, 2000) in which RT is elevated when attention must be redeployed to a region of the environment just explored.

Our experiment was designed to test the efficacy of truth-reversal in detecting deception. Young adults of different ethnicity were asked to lie about particular facts or attitudes during speeded responding either with or without truth reversal. We then determined whether RT (corrected for accuracy) correlated with the veracity of responses in different conditions.

Experiment

Stimuli

We constructed eight *facts*, each with two mutually exclusive and exhaustive *versions*. All 16 statements (8 facts each with 2 versions) are listed in Table 1. During timed procedures, statements were abbreviated slightly; for example, "I write with my left hand" was rendered "write with left hand."

Similarly, we constructed eight *attitudes*, each with two mutually exclusive and exhaustive *versions*. The resulting 16 statements are shown in Table 2.

Participants

Three groups of participants were recruited from the vicinity of the University of Michigan. They were:

(a) 26 Americans of European ancestry (the *European American* group)

(b) 26 East Asians (the *East Asian* group)

(c) 30 native Arabic speakers (the *Arab* group)

Sixteen of the Arabs and all of the European Americans were born in the U.S. The East Asians were born and finished high school in China, Korea, or Taiwan. All the

Arabs were male. The other two groups were equally divided by gender. The average age of the participants (similar across groups) was 23.7, SD = 5.05.

Selecting a lie

Subjects participated individually, performing both the *fact* and the *attitude* procedures (with order counterbalanced across subjects). At the beginning of a given procedure, they rolled an eight-sided die, recorded the number that emerged, and agreed to systematically lie about the corresponding fact or attitude. Next, the subject was presented with the eight facts or attitudes via a computer interface. They marked the version that was true (or "most true") of themselves, taking account of their assigned lie. (That is, for one item the incorrect version was marked as "true.") Subsequently, the subject performed three runs of each of the *uncrossed* and *crossed* procedures in the order uncrossed, crossed, uncrossed, crossed, uncrossed, crossed. Finally, they repeated the foregoing procedure for the remaining kind of item (fact or attitude).

The uncrossed procedure

Each uncrossed run consisted of 32 trials performed via computer monitor and mouse. The trials corresponded to two randomly ordered repetitions of the 16 statements in a given stimulus set (facts or attitudes). A trial began with presentation of one statement visually centered on the screen, using a font size that facilitated normal reading. At the two lower corners were the labels true and false; their left/right positions were determined randomly for each trial. The subject clicked the appropriate label, ending the trial. Trials were preceded by a fixation interval that lasted one second.

The crossed procedure

Crossed runs were identical to uncrossed with one exception. Subjects were required to click the false button for true statements, and click the true button for false statements. Once again, the left/right placement of the labels was randomly determined for each trial. For both uncrossed and crossed trials, responses were supposed to take account of the assigned lie.

Results

We computed the error rate and RT for each fact or attitude by averaging over the two versions of the item. There were thus 12 uncrossed trials and 12 crossed trials corresponding to a given subject and fact, and likewise for attitudes. (12 trials for a given item arise from 3 runs $\times$ 2 repetitions per run $\times$ 2 versions of each item.) These averages are the inputs to the analyses that follow.

RT and accuracy by group

We say that a trial was "accurate" if the response corresponded to what the participant marked as true or false at the beginning of the procedure; otherwise, the trial resulted in "error." Thus, accuracy required clicking on true for the true version of the item in the uncrossed trials, and false for the false version — and the reverse pattern in crossed trials.

Uncrossed and crossed trials were collapsed to compute the average RT for a given fact or attitude, and its error rate. Statistics for the means of these average are shown in Table 3, per group. There is a tendency for the European Americans to be faster than the other groups, and for the East Asians to be more accurate. Welch two sample t-tests yield only the following significant differences. European American responded faster to attitude statements than did East Asians ($t(49) = 2.8$, $p < 0.006$) and Arabs ($t(50) = 3.0$, $p < 0.003$). Also for attitudes, East Asians were more accurate than Arabs ($t(46) = 2.8$, $p < 0.007$).

Comparison of uncrossed versus crossed trials, and lies versus truth

In the Introduction we raised the possibility that taxing working memory via the Stroop-like "crossing" manipulation would have greater impact on lying compared to truth-telling. Using RT as dependent variable reveals little evidence of the desired interaction. This can be seen from Table 4, which records RTs for uncrossed versus crossed trials, split by truth-telling versus deception. Only the East Asians manifest a tendency for the interaction inasmuch as the difference between lying and truth-telling is slightly greater for crossed trials than uncrossed (for both facts and attitudes). An

(unbalanced) factorial ANOVA (uncrossed/crossed, truth/deception) yields insignificant interactions in all cases. Main effects of crossing were significant ($p < 0.05$) only for Arabs and East Asians evaluating facts. The main effect of lying was significant ($p < 0.05$) in all six cases, i.e., greater RT for lying.

When error rate is used as dependent variable, all six data sets exhibit greater difference between truth and deception in the crossed condition compared to uncrossed, consistent with the desired interaction. See Table 5. But this trend does not often reach statistical significance. When the factorial ANOVA used for RT was applied to error rate, the interaction of uncrossed/crossed with truth/deception was significant for European Americans evaluating facts ($p < 0.05$). The interaction is also significant for East Asians judging facts ($p < 0.001$). No other interactions reached significance. All main effects were significant ($p < 0.05$) except for uncrossed/crossed when Arabs evaluated facts, and uncrossed/crossed when European Americans evaluated attitudes.

In sum, compared to truth-telling, lying raised RT and erroneous responding in both uncrossed and crossed trials; for the latter dependent variable, there was a tendency for the increase to be greater for crossed trials. We now attempt to exploit these variables to predict for each subject which of the eight items was chosen for deception. For this purpose, an amalgam of RT and error will prove useful. We define a subject's *performance* on a given item I (fact or attitude in either the crossed or uncrossed conditions) as follows.

$$\text{performance}(I) = \text{RT}(I) \times [1.0 + \text{err}(I)]$$

where $\text{RT}(I)$ is the average RT over the 12 trials involving I, and $\text{err}(I)$ is the percentage of errors the subject made over those trials.

Predicting the deception item

For each participant individually, we attempted to predict which of the eight items was chosen to lie about. The predictions have the form:

Of the eight items, the one lied about has the maximum *index*.

The following seven indexes were defined for each item and participant ($N = 12$ trials in each case).

Uncrossed RT: The mean reaction time for the item during uncrossed trials.
Crossed RT: The mean reaction time for the item during crossed trials.
Uncrossed error rate: The percent of errors for the item during uncrossed trials.
Crossed error rate: The percent of errors for the item during crossed trials.
Uncrossed performance: The mean performance for the item during uncrossed trials.
Crossed performance: The mean performance for the item during crossed trials.
Uncrossed performance + crossed performance: The sum of the two preceding indexes.

Note that by an index being "maximal" for the lie-item, we mean *uniquely maximal*. In other words, if the greatest value of an index was shared by more than one item, the index was not qualified as predicting the lie-item for that participant (even if the lie-item possessed maximal value on that index).

Results for facts are presented in Table 6 by group. For example, in the "Exact predictions" column we see that for 14 of the 26 participants, the lie-item had the greatest uncrossed RT. The probability of such accuracy (or better) by chance is less than 1% by binomial test (the chance of randomly choosing the lie-sentence is 1/8). The most successful index appears to be the last one (uncrossed performance + crossed performance); it predicts the lie-item reliably for the European Americans and East Asians.

Table 6 also shows the number of times an index was highest or second-highest for the lie-item (see the column titled "Top 2 predictions"). Similarly to the exact case, a successful prediction of this kind requires no ties with any of the remaining six items. Top 2 predictions have practical significance because the veracity of one of the two candidate lies might be known on independent grounds. Uncrossed performance + crossed performance is again the most accurate index, predicting reliably for all three groups $p < 0.01$ by a binomial test where accurate random prediction has chance 1/4).

The top of Table 7 displays the accuracy of uncrossed performance + crossed performance by fact, collapsing over the three groups. A majority of the false predictions ("misses")

occurred for the gender and age items (numbers 1 and 4 in Table 1). It might be relatively easy to lie about one's age, possibly because the truth changes so frequently. The same explanation does not apply to gender, however. Overall, uncrossed performance + crossed performance predicted the lie in 38 out of 82 participants.

Results for attitudes are presented in Table 8 by group. As with facts, the best predictor is uncrossed + crossed performance, which accurately identifies the lie in a majority of each group (in 58 of the 82 participants overall). Likewise, the "top 2" predictions are most accurate with this index, Comparison of Tables 6 and 8 reveals greater success in predicting lies about attitudes compared to facts (hit rates of 70.7% and 46.3%, respectively). The bottom of Table 7 displays the accuracy of uncrossed performance + crossed performance by attitude, collapsing over the three groups. Hits and misses are more evenly distributed than for facts. The greatest success in lying (escaping detection) involves attitudes about gun control.

Discussion

For both crossed and uncrossed procedures, we defined the "performance" on a given item to be its average RT multiplied by 1.0 plus its error rate. This simple integration of speed and accuracy allows crossed plus uncrossed performance to predict lying significantly better than chance. Indeed, the latter index was greatest in 46.3% of participants for their deceptive factual item and in 70.7% of participants for their deceptive attitude item. In both cases, the chance baseline (1 in 8 items) is 12.5%. The level of accuracy we achieved required the inclusion of crossed trials, in which responses were generated in the face of Stroop-like truth reversal; uncrossed trials alone were of distinctly less predictive value. Our findings thus validate the strategy of taxing working memory concurrently with response selection. Speed and accuracy are thereby rendered more potent indicators of prevarication.

On the other hand, we did not detect the predicted interaction between lying and truth-telling for crossed versus uncrossed responding; there was merely a trend in this direction using error rate as dependent variable. It thus remains possible that the only advantage of the crossed condition was to add trials to the calculation of the performance index. Doubling the number of uncrossed trials (with no crossed condition

at all) might have allowed equally accurate prediction of the deceptive item. New data are required to decide the issue.

Further investigation is also needed to clarify the role of linguistic variables in RT and error rate. English was not the mother tongue of our East Asian or Arab participants. It would thus be interesting to replicate the experiment within their native languages. In this connection, we note that our attitude stimuli were longer (more words) than the fact stimuli, which likely explains the higher RT for the former items compared to the latter. This difference might not be obtained in other languages. (All our attempts to improve predictions by factoring in stimulus length failed.)

We conclude by emphasizing the limitations of the present study. Most importantly, the stakes for our participants were low (no crime was under investigation). Also, participants had no control over which item was to be lied about (the roll of a die made the choice for them). It is possible that our performance index would be useless in the context of motivated liars with intimate connection to the content of their prevarication. Such people might also deploy countermeasures against our method, notably, very slow responding to all items. The obvious remedy (to explore in future research) would be to require responses within progressively narrower temporal intervals (with error rate as the principle index of deception).

More generally, we do not conceive our paradigm as leading to a "Pinocchio's nose" (a fully reliable indicator of prevarication, Vrij, 2008). Instead, it would best be exploited as one component of a battery of procedures leading to an empirically grounded estimate of the probability of truth-telling versus lying. Even in this limited role, alternative versions of our paradigm need to be developed for populations that differ in literacy and computer use. Beyond RT and error rate, a rich set of dependent variables can be explored in the context of our procedure. These include electrodermal activity and pupil activity (Vendemia et al., 2006; Granholm and Steinhauer, 2004), along with measures of attempted "correction" of responses once given.

Tables

Table 1: Facts used as stimuli

1	I am male	I am female
2	I write with my left hand	I write with my right hand
3	I was born in the USA	I was born in a foreign country
4	I am 20 years old or more	I am 19 years old or younger
5	I have been to Canada	I have never been to Canada
6	I was born in first 6 calendar months	I was born in last 6 calendar months
7	I have seen a Harry Potter movie	I have never seen a Harry Potter movie
8	I rode the subway today	I did not ride the subway today

Table 2: Attitudes used as stimuli

1	There are times that torture is acceptable.	Torture is never acceptable.
2	I believe people should not be allowed to buy guns.	I believe people should be able to buy guns.
3	I believe homosexuality is immoral.	I have no moral problems with homosexuality.
4	I support the Iraq war.	I think the Iraq war was a mistake.
5	I believe the Bush administration has largely been a failure.	I believe the Bush administration has been a success for the most part.
6	I believe abortion should always be illegal.	I believe there are times that abortions should be legal.
7	I believe children are better off when mothers work.	I believe children are better off when mothers stay at home.
8	I think the US threatens the world more than it helps it.	I think the US helps the world more than it threatens it.

Table 3: RT and error rate by group for facts and attitudes, collapsing over uncrossed and crossed trials

Condition, Measure	*Mean*	*SD*	*Median*	*N*
European Americans				
Facts, RT (ms)	2119	622.27	2044	26
Facts, Error (%)	6.89	7.87	4.43	26
Attitudes, RT (ms)	2937	846.31	2765	26
Attitudes, Error (%)	9.11	9.79	5.47	26
Arabs				
Facts, RT (ms)	2382	585.61	2212	30
Facts, Error (%)	7.00	7.49	4.17	30
Attitudes, RT (ms)	3832	1333.6	3386	30
Attitudes, Error (%)	10.23	7.77	7.29	30
East Asians				
Facts, RT (ms)	2381	631.27	2285	26
Facts, Error (%)	3.73	4.66	2.08	26
Attitudes, RT (ms)	3643	946.50	3512	26
Attitudes, Error (%)	5.63	4.16	4.17	26

Table 4: Mean RT (and SD) for lies and truths in uncrossed versus crossed trials, by group

Group	*Uncrossed RT truth*	*Uncrossed RT lie*	*Crossed RT truth*	*Crossed RT lie*
Facts				
European Americans	1818 (521.5)	2301 (761.1)	2304 (1139.3)	2747 (984.5)
Arabs	2059 (768.5)	2477 (995.6	2611 (894.5)	2950 (1441.0)
East Asians	2127 (651.7)	2603 (1453.7)	2500 (842.5)	3113 (1925.5)
Attitudes				
European Americans	2621 (942.1)	3525 (1518.1)	3018 (1145.2)	3992 (1637.1)
Arabs	3396 (1572.2)	4729 (2112.7)	3986 (1802.2)	4902 (1914.6)
East Asians	3306 (1244.6)	4645 (1587.7)	3686 (1198.3)	4697 (2130.6)

Note: Average RTs are rounded to the nearest millisecond. The number of observations for *uncrossed RT truth, European Americans* is 182. This results from 26 subjects times 7 items (out of 8) for which truth was required. The number of observations for *uncrossed RT lie, European Americans* is 26, which results from 26 subjects times 1 items (out of 8) for which deception was required. The number of observations for the other cells of the table are determined similarly.

Table 5: Mean error rate (and SD) for lies and truths in uncrossed versus crossed trials, by group

Group	*Uncrossed RT truth*	*Uncrossed RT lie*	*Crossed RT truth*	*Crossed RT lie*
Facts				
European Americans	3.89 (8.15)	15.4 (26.53)	6.0 (10.95)	25.6 (32.65)
Arabs	4.0 (7.25)	17.2 (30.40	6.5 (11.58)	21.4 (32.29)
East Asians	1.8 (4.04)	4.2 (10.07)	3.8 (7.91)	16.7 (32.66)
Attitudes				
European Americans	6.7 (13.81)	19.6 (25.27)	7.5 (13.71)	26.9 (34.83)
Arabs	6.0 (10.81)	24.7 (30.12)	9.1 (12.84)	32.8 (31.71)
East Asians	2.7 (8.40)	14.4 (18.34)	4.9 (10.66)	22.8 (32.19)

Note: Error rates are given as percentage of mistaken trials. The number of observations for each cell is determined as in Table 4.

Table 6: Predictive accuracy for *facts* of various indexes of item difficulty

Index	*Group*	*N*	*Exact predictions*	*Top 2 predictions*
Uncrossed RT	Eur. Americans	26	14*	17*
	Arabs	30	8	13
	East Asians	26	6	13
Crossed RT	Eur. Americans	26	9	11
	Arabs	30	5	8
	East Asians	26	9	14
Uncrossed error	Eur. Americans	26	8	10
	Arabs	30	7	9
	East Asians	26	3	4
Crossed error	Eur. Americans	26	8	9
	Arabs	30	5	11
	East Asians	26	7	7
Uncrossed performance	Eur. Americans	26	14*	19*
	Arabs	30	11	17*
	East Asians	26	5	12
Crossed performance	Eur. Americans	26	12*	15*
	Arabs	30	10	16*
	East Asians	26	12*	16*
Uncr + Cr performance	Eur. Americans	26	15*	19*
	Arabs	30	11	17*
	East Asians	26	12*	17*

Note: N records the number of participants in each group. An "exact prediction" occurs when the index is greatest for the lie item (1 of 8). A "top 2 prediction" occurs when the index is either greatest or second-greatest for the lie item (2 of 8).

*Starred results have low probability ($p < 0.01$) of arising from independent trials with probability of accurate lie-spotting 1/8 or 1/4 (for exact and top 2 predictions, respectively).

Table 7: Predictive accuracy of crossed + uncrossed performance by item (facts and attitudes)

Facts:	1	2	3	4	5	6	7	8	Tot.
Hits:	2	6	2	3	6	5	7	7	38
Misses:	14	5	4	11	3	3	3	1	44
Sum:	16	11	6	14	9	8	10	8	82

Attitudes:	1	2	3	4	5	6	7	8	Tot.
Hits:	4	5	10	9	6	10	4	10	58
Misses:	4	9	3	1	3	2	2	0	24
Sum:	8	14	13	10	9	12	6	10	82

Note: For facts and attitudes separately, we show the predictive success of the index crossed performance + uncrossed performance. See Tables 1 and 2 for interpretation of item numbers.

Table 8: Predictive accuracy for *attitudes* of various indexes of item difficulty

Index	*Group*	*N*	*Exact predictions*	*Top 2 predictions*
Uncrossed RT	Eur. Americans	26	12*	16*
	Arabs	30	14*	19*
	East Asians	26	13*	16*
Crossed RT	Eur. Americans	26	8	16*
	Arabs	30	14*	17*
	East Asians	26	10*	14
Uncrossed error	Eur. Americans	26	10*	11
	Arabs	30	11*	15
	East Asians	26	12*	12
Crossed error	Eur. Americans	26	11*	12
	Arabs	30	14*	20*
	East Asians	26	10*	11
Uncrossed performance	Eur. Americans	26	13*	18*
	Arabs	30	17*	23*
	East Asians	26	18*	19*
Crossed performance	Eur. Americans	26	11*	19*
	Arabs	30	18*	22*
	East Asians	26	12*	15*
Uncr + Cr performance	Eur. Americans	26	17*	20*
	Arabs	30	21*	25*
	East Asians	26	20*	21*

Note: N records the number of participants in each group. An "exact prediction" occurs when the index is greatest for the lie item (1 of 8). A "top 2 prediction" occurs when the index is either greatest or second-greatest for the lie item (2 of 8).

*Starred results have low probability ($p < 0.01$) of arising from independent trials with probability of accurate lie-spotting 1/8 or 1/4 (for exact and top 2 predictions, respectively).

References

E. Granholm and S.R̃. Steinhauer, editors. *Pupillometric measures of cognitive and emotional processes*, volume 52. International Journal of Psychophysiology, 2004.

R. M. Klein. Inhibition of return. *Trends in Cognitive Sciences*, 4(4):138–147, 2000.

A. W. MacDonald, J. D. Cohen, V. A. Stenger, and C. S. Carter. Dissociating the role of the dorsolateral prefrontal and anterior cingulate cortex in cognitive control. *Science*, 288(9): 1835–1838, June 2000.

C. M. MacLeod. Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin*, 109(2):163–203, 1991.

J. V. Pardo, P. J. Pardo, K. W. Janer, and M. E. Raichle. The anterior cingulate cortex mediates processing selection in the Stroop attentional conflict paradigm. *Proc. Natl. Acad. Sci.*, 87(1):256–259, 1990.

S. L. Sporer and B. Schwandt. Paraverbal indicators of deception: A meta-analytic synthesis. *Appl. Cognit. Psychol.*, 20:421446, 2006.

J. R. Stroop. Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, 18:643–662, 1935.

J.M̃. Vendemia, M. J. Schilliaci, R. F. Buzan, E. P. Green, and S. W. Meek. Credibility assessment: psychophysiology and policy in the detection of deception. *American Journal of Forensic Psychology*, 24(4):53–85, 2006.

A. Vrij. *Detecting Lies and Deceit: Pitfalls and Opportunities*. John Wiley & Sons, West Sussex UK, 2nd edition, 2008.

M. Zuckerman, B. M. DePaulo, and R. Rosenthal. Verbal and nonverbal communication of deception. In L. Berkowitz, editor, *Advances in Experimental Social Psychology*, volume 14, pages 1–57. Academic Press, New York NY, 1981.