Do Liars Blink Differently? Automated Blink Detection during Deceptive Interviews

Aaron C. Elkins
Imperial College London
a.elkins@imperial.ac.uk

Nikolaos Sorros
Imperial College London
nikolaos.sorros12@imperial.ac.uk

Stefanos Zafeiriou
Imperial College London
s.zafeiriou@imperial.ac.uk

Judee K. Burgoon
University of Arizona
jburgoon@cmi.arizona.edu

Maja Pantic
Imperial College London
m.pantic@imperial.ac.uk

Jay F. Nunamaker
University of Arizona
jnunamaker@cmi.arizona.edu

Abstract

This research investigates the development and use of an automated blink detection algorithm applied to a large (N=176) deceptive interview video corpus. The automated blink detection algorithm was 93% accurate. This work represents the first analysis of deceptive blinks of this magnitude (46 hours of video) and degree of ecological validity. After applying the algorithm to the interview video corpus, deceivers were found to blink less when lying to cognitively demanding questions. In addition to deception, people blinked more over time, more when older, and less when more skilled in social expressivity. The results of this study suggest that any deception detection algorithm that relies on blinks needs to account for time, interviewee demographics and social skill, question type, and turn taking phase (interviewee listening or speaking).

1. Introduction

The eyes have long been considered a window into our inner thoughts and intentions. One long standing behavioral stereotype is that when someone averts their gaze they are being dishonest or are feeling ashamed. Despite the myriad cultural and interactional reasons for averted gaze, this belief has pervaded popular culture.

Gaze aversion has been proposed to reflect the mind constructing ideas (up and to the right) or remembering (up and to the left) by adherents of Neuro-linguistic programming. Accepting this simple mind-eye gaze relationship ignores the fact that both liars and truthtellers engage in constructing and remembering. The growing evidence (example: [1], [2]) discrediting gaze aversion as an indicator of deception has not impeded its belief or use in practice.

While incredulity is well warranted anytime a simple heuristic such as gaze aversion is employed, this should not be interpreted as meaning eye behavior such as blinks, gaze, fixations are not diagnostic for detection deception.

2. Blinks

We blink often throughout the day. Only a small fraction of our blinking provides lubrication of the eyes. After our eyes are properly lubricated, what function do blinks serve? When we blink, we close off the visual channel and disengage from any distracting visual stimulus and reallocate cognitive resources for concentration [3], [4].

Conversely, we inhibit our blinks to maximize the gathering of visual stimulus and minimize any potential information loss [5]. One need only notice how we have the tendency to inhibit our blinks until the end of a sentence when reading to observe this phenomenon in action.

2.1. Blinks and Deception

The relationship between blinks and communication, both truthful and deceptive are related to the cognitive demands and mental vigilance of the speaker. Deception is predicted to be more cognitively demanding, particularly when the responses require more complexity and are under scrutiny [6], [7]. This implies that different questions

will elicit different patterns of blinks as a function of their cognitive complexity.

One challenge to understanding the complex relationship between blinks and deception is the immense degree of labor required to measure eye behavior, which is primarily done manually. This limitation has artificially reduced the number of studies, participants in the studies, and the length/realism of the experimental procedures.

In an effort to mitigate the eye measurement problem, Leal and Vrij [8] attached electrodes to the participants to measure activation about the eye. Similarly, Fukuda investigated an automated video analysis of blinks during deception, but it required a helmet mounted camera [5]. Both of this studies have a maximum of 26 participants and employ a fixed phase non-interactive protocol. Fukuda used a modified guilty knowledge paradigm where deceivers named playing cards they were shown over multiple trials.

These studies provide some evidence for the relationship to deception, but there is still a great need to analyze eye behavior during realistic interactions.

To address this, we developed an automated method of measuring blinks from video. We then applied this method to a large video corpus of 176 interviews where subjects were randomly instructed to lie to questions. The interviews lasted an average of 16 minutes and were interactive (i.e., interviewer rated their credibility and asked follow-up questions.) To account for the complexity of responses, different question types, differing on cognitive demands and verifiability were included.

In total, over 46 hours of video interviews were analyzed using the blink detection method discussed next. This work represents the first analysis of deceptive blinks of this magnitude and degree of ecological validity.

3. Automated Blink Detection Methods

3.1. Optical Flow

Optical flow techniques have been used extensively in blink detection. Optical flow captures the relative movement of pixels during time. This technique has been used for both eye localization and feature extraction [9], [10]. In eye localization optical flow takes advantage of the fact that blinks occur frequently thus frequent movement in a specific area is translated into a bounding box. For feature extraction, the researchers rely on the fact that blinks produce a distinct downward and upward movement that can characterize the blink phenomenon. The

drawback of these methods is that they rely on a far from perfect technique which is that of optical flow. The best accuracy achieved is 97%.

3.2. Histogram of Oriented Gradients (HOG)

Histogram of Oriented Gradients (HOG) is a state of the art technique for object recognition. HOG works by calculating the histogram of oriented gradients of the image. The votes of the histogram are the gradient magnitudes and the number of bins (typically 9 bins) representing a spectrum of [0,180] degrees. These histograms are calculated after the image has been split into blocks of usual size 8x8 pixels. HOGS in blink detection are used to identify open and close eyes states and require and eye localization step first. The best accuracy reported so far is 90.28% [11].

3.3. Gabor filters

Gabor filter is another descriptor that can be used instead of HOG to identify open and close eye states and as HOG they need an eye localization pre-step. The best performance reported is 93% [12].

3.4. Hidden Markov Models (HMM)

Hidden Markov models have been used to describe the temporal dimension of blinks. HMM is a statistical Markov model for which the system contains unobserved hidden states. In blink detection HMM were used to capture the transition from open to half open to closed eyes and back. This method relies on an accurate eye localization step. The accuracy reported in the only attempt we came across in the literature was 91% [13].

4. Proposed Method

All methods mentioned above depend on accurately locating the eyes. This is a far from trivial problem in computer vision. A more general problem is that of facial feature detection. Most common techniques align to one of three categories: Parameterized appearance models (e.g. AAM), Discriminative approaches, and Part-based Performable models.

The general pipeline that the blink detection techniques we discussed follow, is that of locating an area that contains the eyes and then somehow extracting information, features, relevant to the blinking.

Our approach is fundamentally different. Instead of separating the steps for eye localization and feature extraction we have one step that does both. Our method is based in a facial feature detection technique recently presented in CVPR [14]. The method is called Supervised Descent (SDM) and is presented below.

4.1. Supervised Decent Method

Facial feature detection techniques usually involve a model and an optimization technique. Advances in the field usually come from introducing a new model or altering an existing one. SDM focuses on the optimization step and introduces a better optimization technique tailored for face alignment applications.

Let I∈R^n be a face image of n pixels, I(x)∈R^p represent p landmarks on a face, F a non-linear function for constructing feature descriptors (if it is SIFT features, which have 128 dimensional descriptor, then F(I(x))∈R^128p. Manually labeled landmarks are denoted as x^*, it should be noted that they are known during the training phase. From all of the available training samples mean landmarks are calculated, initialized using a face detector and denoted as x_0. Given this set up, facial feature detection task can be formulated as an optimization problem of the objective function [14].

$$f(x_0 + \Delta x) = \| F(I(x_0 + \Delta x)) - F(I(x_*)) \|_2^2 \quad (1)$$

Newton method is one the most common techniques used in optimization. Let's assume that our function is twice differentiable for now, an assumption we will drop later, so that we can apply Newton.

$$x_{k+1} = x_k - H^{-1} J_f \quad (2)$$

Newton method is essentially searching for the local minimum by following the descent direction $H^{-1} J_f$.

Now using the chain rule $J_f = 2 J_F^T (\varphi_0 - \varphi_*)$, where $\varphi(x) = F(I(x))$ we can rewrite Newton method as: $\Delta x = R_0 \varphi_0 + b_0$ where b models x* which is unknown.

Calculating the matrix R is not always feasible. H is not always invertible and even when it is the time complexity of the inversion (O(n3)) makes it a bad choice. To cope with that a lot of methods use some sort of approximation of Hessian (Levenberg–Marquardt modification, Gauss-Newton, Quasi Newton etc). Unfortunately these approximations drop the accuracy of the methods significantly.

Instead of using another layer of approximation, SDM decides to perform a step-wise linear regression to learn generic gradient directions, matrices R and b.

$$x_k = x_{k-1} + R_{k-1} \varphi_{\kappa-1} + b_{k-1} \quad (3)$$

The difference with Newton method is that Newton generates image specific descent directions whereas SDM generates generic ones.

To perform regression we need to define a cost function, a step which is straightforward at this point.

$$\arg \min_{R_o b_0} \sum_{I^i} \int p(x_0^i) \| \Delta x^i - R_0 \varphi_0^i - b_0 \|^2 \, dx_0^i \quad (4)$$

Where $\Delta x^i = x_*^i - x_0^i$, $\{I^i\}$ is a set of images. The initial point comes from a face detector, inaccuracy of which is modeled by sampling x_0 from a normal distribution. We will use Monte Carlo numerical integration to estimate the integral.

$$\arg \min_{R_o b_0} \sum_{I^i} \sum_{x_0^i} \| \Delta x^i - R_0 \varphi_0^i - b_0 \|^2 \quad (5)$$

This is a least square problem that has well known closed form solution.

$$\widetilde{R_0} = \sum_{d_i} \sum_{x_0^i} \Delta x_i \tilde{\varphi}_0^T \left(\sum_{d_i} \sum_{x_0^i} \tilde{\varphi}_0^T \tilde{\varphi}_0 \right)^{-1} \quad (6)$$
where $\tilde{R}_0 = [R_0, b_0]$ and $\tilde{\varphi}_0 = [\varphi_0; 1]$

The novelty of this technique is the step-wise regression unlike other attempts that experimented with one step. This means that we calculate different descent directions for each step. The algorithm seemed to always converge after 5 steps.

4.2. Normalization

Before we introduce blinks, we need to cope with the pose and scale variance. To do that we need to define the shape s= [x1,y1,...,xp,yp] where x and y come from the points we tracked or from the ground truth. Removing pose and scale similarities is as easy as projecting into the eigenvectors that contain pose and scale information and remove those from the shape. Pose eigenvectors can be obtained by performing PCA and empirically watching how many of the first 6 eigenvectors contain pose information. For the scaling we will use eigenvectors coming from the mean shape.

$$s^{free} = s - P(P^T(s - \bar{s})) + \bar{s} \quad (7)$$

Where $P \in \mathbb{R}^{px4}$, is a matrix containing the 4 first eigenvectors from the PCA analysis, that explain

pose variation in the dataset. A similar formula removes scale variations as well.

Figure 1. Supervised Decent Method Facial Feature Tracking Applied to Deceptive Interview Videos

5. Blink detection

Having accurately calculated and normalized the points around the eyes we can extract information about both blink count and duration. Starting with blink count we can measure the root mean distance from the upper eye points to the lower ones.

$$distance(i) = \sqrt{\frac{1}{4}(x_i^d - x_i^u)^T(x_i^d - x_i^u)} \qquad (8)$$

where $x_i^d, x_i^u \in \mathbb{R}^4$ are vectors representing the downer {d} and upper {u} points in time i.

5.1. Validating Threshold

To examine the optimal threshold for identifying a blink we created a corpus of 3,000 blinks and non-blink examples from 30 random subjects. Our cost function is to optimize the minimum classification error.

5.2. Accuracy

Next the corpus was split into a training and independent test set. The thresholds were validated and trained on the training set and applied to the test set. The final overall accuracy of the blink detector was 93%.

Figure 2. Blink Detector Applied to Deceptive Interview Corpus

6. Deceptive Interview Experiment

After validating the accuracy of the blink detector, it was applied to the deceptive interview corpus to measure blinks. The corpus contains 201 multicultural interviews where interviewees completed a 24 questions interview (12 short and 12 long response questions). For a more thorough discussion of the experimental method see [15], [16].

During the interview, participants were randomly assigned to lie or tell the truth to the questions asked of them by professional interviewers. The interviewers rated how honest they found the interviewees. The participants received financial compensation based on how credible they were found to be.

The questions asked of the subject, detailed in Table 1, varied on how cognitively demanding they were to respond to. The questions varied on cognitive judgment and verifiability, discussed next.

6.1. Sample

Of the 201 videos, 25 resulted in tracking errors during the automated blink analysis and were omitted from analysis. The final analyzed dataset contained 176 deceptive interviews of which 54.5% were male and the average age was 30 (SD = 14).

6.2. Question Types

6.2.1. Question Cognitive Judgment

Cognitive judgment is classified as factual, divergent, or evaluative. Factual questions are designed to elicit simple, straightforward answers at the lowest level of cognitive judgment. Answers to these questions may be based on known facts or history and require simple recall or logic. Divergent questions require a higher level of cognitive judgment, as they may require prediction, conjecture, or inference, based on facts. Answers to these

questions may vary significantly, as they often require creative thought. Evaluative questions require the highest level of cognitive judgment and may require synthesis, evaluation, and sophisticated conclusions. A truthful response would be quickly produced to factual questions but require more time to respond to evaluative and divergent questions. Two of the questions (Question 3 and Question 10) were specifically what would be labeled as control or comparison questions in polygraph examinations. They were intended to elicit arousal due to being conflicted about admitting to immoral or unethical conduct. As they were expected to create more ambivalence for truthful than deceptive respondents, they were expected to require more time and create more cognitive difficulty for truthful than deceptive responses.

6.2.2. Verifiability

Question verifiability is classified as verifiable, not verifiable. Questions that are verifiable may be verified as true or false by the interviewer, while questions that are not verifiable may not be. A truthful interviewee would not take into account the verifiability of the question. The more verifiable the answer, the more careful the deceiver should be in crafting a convincing answer.

Table 1. Question and Types

Questions	Cognitive Judgment	Verifiable
1. What is the worst job you ever had and why did you dislike it?	Evaluative	No
2. What would you do if your boss gave you credit for someone else's work?	Divergent	No
3. Tell me about a time when you thought of stealing something valuable from someone.	Factual	No
4. Please tell me everything you did today from leaving your home to arriving for this interview.	Factual	Yes
5. What do you consider to be your greatest strengths?	Evaluative	No
6. What else are you going to do today? Who will you see and where will you go?	Factual	Yes
7. Think about people that really irritate you. Why do they bother or annoy you?	Evaluative	No
8. What do you plan to do during your next break or vacation?	Factual	Yes
9. Remember the room where you arrived for the experiment. Tell me everything about that room and what happened while you were there.	Factual	Yes
10. Tell me about a time when you told a serious lie to get out of trouble.	Factual	No
11. If you found a wallet containing $1,000 and no identification in it,	Divergent	No
what would you do with it and why?		
12. What is the worst restaurant you ever went to? Why did you dislike it?	Evaluative	No

6.3. Self-Report Measures

Prior to the interview, participants completed a pre-survey to collect demographics (i.e., age and gender) and behavior moderator measurements including social skills and motivation.

6.3.1. Social Skills

Participant social skills and deception skills. Participants can vary substantially in their ability to communicate effectively and lie successfully. To assess the extent of variability in the sample and to incorporate, if necessary, a covariate of social skills, participants were asked to complete an abbreviated version of Riggio's [17] Social Skills Inventory. The full SSI consists of 120 items that measure six dimensions. Three are related to verbal skills. Social expressivity taps into a person's ability to express oneself effectively in a variety of social contexts, social control taps into one's ability to manage and adapt one's verbal communication, and social sensitivity taps into one's ability to interpret others' verbal communication. The other three dimensions concern nonverbal skill. Emotional expressivity reflects one's ability to make one's emotional expressions understood, emotional control reflects the ability to mask one's true feelings, and emotional sensitivity reflects one's ability to read others' emotional states. The dimension of interest in this study was social expressivity.

6.3.2. Motivation

If participants are unmotivated during an interaction, their deceptive and truthful performances will not be representative of what occurs outside the laboratory. A motivation measurement was included asking participants how important it was for them to appear credible.

6.4. Blink Processing

Each of the videos were processed using the previously describe blink detection algorithm. This resulted in a binary blink or no blink indicator for every frame of video. The videos were 30 frames per second. To facilitate analysis, the frame level data were windowed into five second intervals and blink count, the primary dependent measurement, was aggregated within each window. If an interview was 15 minutes long, this would result in 180 time points.

The blink counts were then converted to blinks per minute to aid interpretability.

7. Results

7.1. Blinks and Deception

To assess the relationship between blinks per minute and time, a multilevel growth model was specified with blinks per minute as the response variable (N= 32,613) regressed on a dummy coded variable indicating if the participant was lying. To reflect the repeated-measures experimental design over time, the Intercept of blinks per minute were modeled to vary within Subject (N=176) as random effects.

To test if including the lying variable improved the fit of the model to the data, it was compared against the unconditional means model, which omits any fixed effects using deviance-based hypothesis tests. The difference in deviance statistics was $\chi2(1)$ =7.55 and significant at the p <.01 level. This allows us to reject the null hypothesis that deception does not affect blinks per minute.

When lying, participants increased their blink rate to M=32.5 blinks per minute (SD=27.73) from M=31.81 (SD=26.74). Liars blinked more possibly to allocate more cognitive resources to facilitate the increased demands when responding.

While this result provides some insight into the eye behavior patterns of liars, it still would not provide enough discrimination to reliably use blinks to predict deception. Next, we specified a model containing all of the variables predicted to affect blinks per minute in addition to deception.

7.2. Blinks, Question Type, and Behavior Moderators

A multilevel growth model was specified with blinks per minute as the response variable (N=32,041) regressed on time (in minutes), dummy coded lie variable, Question Cognitive Judgment (coded to contrast Factual as the baseline, the lowest cognitive demand question type), Age (mean centered), Social Expressivity (standardized), Male dummy code, Question Verifiable dummy code, and the participant's reported motivation (standardized). Two interactions were included in the model, Lie x Cognitive Judgment and Lie x Question Verifiability. The Intercept of blinks per minute were again modeled to vary within Subject (N=173) as random effects. Three subjects were omitted because they had incomplete cases.

The complete model is listed below in Table 2. Examining the model estimates we find that the average blinks per minute for someone of an average age answering honesty to a factual question was 31. However, when lying they increased they blinks per minute by 1.67.

All participants increased their blinks per minute over time by .20 a minute. This may reflect their acclimation to the interview where they may have been inhibiting their blinks early on to focus instructions or interviewer behavior cues.

For every year over the average interviewee age, blinks per minute increase by .27 and every standard deviation increase in Social Expressivity resulted in a reduced blinks per minute of 2.20.

Question verifiability had no significant effect on blinks per minute and motivation and gender had a suggestive, but non-significant effect.

Finally, when contrasting the high question type blinks per minute x Lie interaction against the factual question type baseline we find that liars reduce their blinks per minute by 2.38. This suggests that lying to factual questions resulted in more blinks than evaluative questions.

Table 2. Multilevel Growth Model Predicting Blinks per Minute

		Estimate
Initial Status Rate of Change	Intercept	31.03***
	Time (Minutes)	0.20**
	Lie	1.67***
	CogJudge Divergant	-0.52
	CogJudge Evaluative	2.58***
	Age	0.27*
	Social Expressivity	-2.20*
	Male	-3.59~
	Verifiability	0.56
	Motivation	-1.71~
	Lie:CogJudge Divergent	-0.70
	Lie*CogJudge Evaluative	-2.38***
Random Effects - Variance Components (Standard Deviation)		
Level-1:	Within-Subject	23.48
Level-2:	In initial status	12.79
	Deviance	293,862
	AIC	293,889
	BIC	294,007

~p<.10; *p<.05; **p<.01; ***p<.001.

8. Conclusion

This work represents a first and important step in disentangling the complex relationship between eye behavior and deception. One limitation of this work is that the questions were not coded for phases of turn taking (interviewer speaking or interviewee speaking). This has important implications for the pattern of eye behavior. For example, we would predict that a deceiver would be more likely to inhibit their blinking when listening to the interviewer in an attempt to evaluate any visual cues to suspicion. Similarly, we would expect to see more blinking during responding while liars are crafting and delivering their message.

The next step of this research will be to code the specific phases of the interview and examine the more advanced patterns of blinks, more suitable for eventual deception detection.

The results of this study suggest that any deception detection algorithm that relies on blinks needs to account for time, interviewee demographics and social skill, question type, and turn taking phase.

9. References

[1]   R. Wiseman, C. Watt, L. Ten Brinke, and S. Porter, "The eyes don't have it: lie detection and neuro-linguistic programming," *PLoS One*, 2012.

[2]   A. C. Elkins, D. C. Derrick, and M. Gariup, "The Voice and Eye Gaze Behavior of an Imposter: Automated Interviewing and Detection for Rapid Screening at the Border," in *Conference of the European Chapter of the Association for Computational Linguistics*, 2012.

[3]   D. E. Irwin and L. E. Thomas, "Eyeblinks and Cognition," in in *Tutorials in Visual Cognition*, 2010, pp. 121–141.

[4]   T. Nakano, M. Kato, Y. Morito, S. Itoi, and S. Kitazawa, "Blink-related momentary activation of the default mode network while viewing videos," *Proc. Natl. Acad. Sci.*, pp. 3–7, 2012.

[5]   K. Fukuda, "Eye blinks: new indices for the detection of deception," *Int. J. Psychophysiol.*, vol. 40, no. 3, pp. 239–245, 2001.

[6]   D. B. Buller and J. K. Burgoon, "Interpersonal Deception Theory," *Commun. Theory*, vol. 6, no. 3, pp. 203–242, Aug. 1996.

[7]   A. Vrij, S. Mann, R. Fisher, S. Leal, R. Milne, and R. Bull, "Increasing cognitive load to facilitate lie detection: The benefit of recalling an event in reverse order," *Law Hum. Behav.*, vol. 32, pp. 253–265, 2008.

[8]   S. Leal and A. Vrij, "Blinking during and after lying," *J. Nonverbal Behav.*, vol. 32, no. 4, pp. 187–194, 2008.

[9]   R. Heishman and Z. Duric, "Using Image Flow to Detect Eye Blinks in Color Videos," *2007 IEEE Work. Appl. Comput. Vis. (WACV '07)*, pp. 52–52, Feb. 2007.

[10]   T. N. Bhaskar, F. F. T. Keat, S. Ranganath, and Y. V Venkatesh, "Blink detection and eye tracking for eye localization," in *TENCON 2003. Conference on Convergent Technologies for the Asia-Pacific Region*, 2003, vol. 2, no. iii, pp. 821–824 Vol.2.

[11]   K. Minkov, S. Zafeiriou, and M. Pantic, "A comparison of different features for automatic eye blinking detection with an application to analysis of deceptive behavior," *Commun. Control Signal Process.*, no. May, pp. 1–4, 2012.

[12]   J.-W. Li, "Eye blink detection based on multiple Gabor response waves," in *2008 International Conference on Machine Learning and Cybernetics*, 2008, vol. 5, no. July, pp. 2852–2856.

[13]   Y. Sun, S. Zafeiriou, and M. Pantic, "A Hybrid System for On-line Blink Detection," in *Forty-Sixth Annual Hawaii International Conference on System Sciences*, 2013.

[14]   Xuehan-Xiong and F. De la Torre, "Supervised Descent Method and its Application to Face Alignment," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2013.

[15]   A. C. Elkins, "Vocalic Markers of Deception and Cognitive Dissonance for Automated Emotion Detection Systems," University of Arizona, 2011.

[16]   J. K. Burgoon, T. Levine, J. F. J. Nunamaker, D. Metaxas, and H. S. Park, "Rapid Noncontact Credibility Assessment: Credibility Assessment Research Initiative Final Report," 2009.

[17]   R. Riggio, "Assessment of basic social skills," *J. Pers. Soc. Psychol.*, vol. 51, pp. 649–660, 1986.