

Calibration Methods of Deception Detection

**Claudia Castro, Ernesto Octavio Lopez, Guadalupe Elizabeth Morales,
Guadalupe Villarreal**

Cognitive Science Laboratory, Department of Psychology, Universidad Autonoma de Nuevo Leon, Monterrey, Mexico

Email: psiclaudiacaastro@gmail.com

Received 14 September 2014; revised 12 October 2014; accepted 7 November 2014

Copyright © 2014 by authors and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

A sample of judges with different ages (children, young adults and adults) as well as a sample of actors (young adults) was required to participate in a deception detection study. Judges were required to evaluate 16 videos where a person might be lying or not lying about a video content. The study sought to look over three aspects of judges' accuracy judgments related to deception detection (discrimination, calibration and global error) by using calibration graphs. Results showed that some children outperformed adults by better estimating the probabilities of being deceived but they performed the same as both adult groups at discriminating those actors who lied from those who did not lie. It is argued that since children have not been sufficiently exposed to cultural factors related to deceiving behavior, they have better calibration judgment. Implications to deception research are discussed in the paper.

Keywords

Deception Detection, Calibration Graphs, Age and Deception Detection

1. Introduction

A considerable amount of academic reports emphasize our low capacity to detect someone who is lying to us. It seems like if we were born to be deceived and human history was plenty of relevant anecdotes of how this incapacity to catch a liar molded the evolution of our society (e.g. kings tricked to be poisoned or Government leaders led into war by deceitful peace agreements (Troville, 1939). Meta-analytical findings seem to confirm this human inability to detect deception (Bond & DePaulo, 2006) and it is suggested that this poor performance is due to the deceived person's bias to believe that deceivers are prone to tell the truth (Levine, Feeley, McCornack, Harms, & Hughes, 2005).

Even when not all deceptive behavior is harmful, its historical relevance to the evolution of our society as

well as its relevance to our own lives, catching a liar is primordial to survival and social adaptation. Early historical records on how to catch a liar went back up to 300 - 250 B.C. with Erasistratus (Greek physician) who suggested a set of physical signs revealing deception (Caso, Gnisci, Vrij, & Mann, 2005). Since then a great deal of deception detection methods have been tried and these encompass a spectrum that goes from old torture methods (e.g. Ordeal method), witness testimonies, etc. to modern polygraphy (Senter & Dollins, 2003), brain activity records or considering micro face expressions (Fu Ch, Williams, Brammer, Suckling, Kim, Cleare, Walsh, Mitterschiffthaler, Andrew, Pich, & Bullmore, 2007). Furthermore, academic research trends on deception have focused on specific modalities of deception detection and deception production such as facial expression, voice tone, body gestures, etc. (DePaulo, Lindsay, Malone, Muhlenbruck, Charlton, & Cooper, 2003) through integral body expression (Farquhar, 2005; Burgoon, Stoner, Bonito, & Dunbar, 2003). For instance, it has been suggested that tone of voice and cadence of speech allow by themselves a clear footprint for deception detection in humans (Hancock, Thom-Santelli, & Ritchie, 2004; Hauch, Blandón-Gitlin, Massip, & Sporer, 2012) as well as for mechanical lie detectors (Elkins, Derrick, & Gariup, 2012).

Heated debates over the accuracy and limitations of these approaches to deception detection (for a review, see Gokhman, S., Hancock, J., Prabhu, P., Ott, & Cardie, 2012) have suggested looking for new empirical directions to provide insights on how people judges someone who is lying. Here, cognitive approaches to deception detection propose that by understanding cognitive participation during deception detection we can determine specific mental articulation that lead us to accurately determine deceptive behavior and at the same time we empower researchers to understand limitations of humans as lie detectors. For instance, it is suggested that short term memory limitations (like cognitive load; Elfenbein & Ambady, 2002; Blandón-Gitlin, Pezdek, Lindsay, & Hagan, 2009) or activated stereotypes in long term memory based on content analysis of information manipulated by the deceiver (Rogers, Boals, & Drogin, 2011) or using systematic thinking (cognitive algebra) to perceive deceiver's attributes (Castro, Morales, & Lopez, 2012) are factors directly related to human deception detection accuracy.

The current study proposes that even when cognitive determination of cognitive processing underlying judgment to deception detection is fundamental to theory, there is still need for determination of cognitive processing parameters related to accuracy. Specifically, we propose that by extrapolating previous cognitive techniques of judgments accuracy (like calibration graphs e.g. Yates, 1990) into the exploration of how people unsuccessfully or successfully catch a liar, new cognitive measurement can be introduced not only to cognitive modelling but complementary help to benefit traditional detection deception instrumentation. By doing so, a functional cognitive approach is introduced to explore how accuracy cognitive processing parameters like calibration and discrimination (Yates, 1990) are assumed to vary across age typifying our capacity to detect deceiving behavior (Rotenberg & Sullivan, 2003; Sweeney & Ceci, 2014). In order to explore this assumption, the following deception detection study was carried out.

2. Method

The current study constitutes a functional cognitive approach to study deception detection. As far as the authors know after an academic review on current digital bases (EBSCO, PROQUEST, MEDPUB and others), this is the first time that judgment calibration analysis is used to deception detection. As we will argue in the discussion, rather than considering this research as a formal study on age differences and detection deception detection it must be considered as new exploratory empirical direction to widen our knowledge about cognitive functioning underlying our ability to catch a liar.

2.1. Participants

This study consisted of two kinds of participants. First a sample of actors whose intention could be to lie or not lie about a video content. Second a sample of judges of different ages to detect possible deceivers.

Actors were 16 middle class typical bachelor psychology students (randomly recruited) from a city at the North of Mexico whose age ranged between 17 and 23 years old ($M = 20$, $SD = 2.13$). Their participation was voluntary with no monetary reward. Signed consent was required from them. Judges to deception detection consisted of a sample containing three different age groups. The first group consisted of 8 children whose age varied between 7 and 12 years old ($M = 9.25$, $SD = 1.66$). The teen sample consisted of 13 young adults whose age ranged between 18 and 22 years old ($M = 20$, $SD = 1.60$). Finally, a group of 8 adults whose age ranged between 37 and 42 years old ($M = 39.12$, $SD = 2.03$) participated in this study. All judges were randomly recruited from the same cultural context.

2.2. Instrument and Stimuli

In this paper, judgment calibration is understood as a mathematical technique to numerically and visually express how accurate is a person to judge an event outcome. In this case, visualization of probability judgment values indicating deception detection accuracy. **Figure 1** shows instances of calibration graphs. Each graph considers two main aspects of judgment. First over the horizontal axes, how often a specific probability judgment category is used (e.g. 20% or 90% of certainty someone is lying) and second over the vertical axis the proportion of times the target event happened when a judgment category was used. Here, accuracy is assumed to be composed of two identifiable properties. For example, the left panel from **Figure 1** illustrates perfect calibration and discrimination to 10 judgments (the number on the dots tells the number of times the category was selected) whereas the middle panel shows perfect calibration and not good discrimination. The closest the points fall over the line the better the calibration score is. On the other hand the closest the dots fall to the extremes of the axes best discrimination is obtained.

Numerically speaking, the closest to zero a Calibration Index value is (*CI*) then better judgment calibration is obtained ($0 \leq CI \leq 1$). On the contrary the more distant a Discrimination Index value (*DI*) gets from zero then better discrimination is obtained. Thus human judgment accuracy is assumed to be composed at least with these two components. Furthermore, a Probability Score value (*PS*) can be obtained by following this analytical approach. Thus, *PS* is a numerical indicator of how far is a person from being a perfect judge or in other terms a perfect human deception detector. Mean *PS* values between zero and 0.50 typify a variety of judges such as being a clairvoyant who knows all the correct answers if this value is zero or an expert judge described by a 0.17 mean *PS* score.

As an example let us consider the calibration graph shown on the right panel from **Figure 1**. Here a judge had to provide probability judgments about if someone was lying regarding a video content. **Table 1** shows 25 deception detection probability values of someone telling the truth or not about a video content. Here, the obtained judge's accuracy judgments values to deception detection were $PS = 0.1476$, $CI = 0.0036$, $DI = 0.1018$. These values are obtained by first subtracting outcome index values (fourth column) from probability judgments (second column). Then squared differences of each value are summed to obtain the mean probability score. As pointed before, this value indicates how far we are from someone who never fails. Note that the third column (from left to right) indicates the true intention of a possible deceiver whereas the fourth column is a numerical representation of intention (outcome index). The sample base rate is the proportion of times the event occurred (pointed by a horizontal dotted line on the right panel from **Figure 1**) and is the mean of outcome index values (d). Regarding discrimination, perfect discrimination is indicated by each of the points in a calibration graph falling either at the top of the graph or at the bottom part whereas nil discrimination is indicated by dots falling over the sample base dotted line. Formal computation to obtain discrimination and calibration values can be observed in **Table 2** (for a complete review see [Yates, 1990](#)).

CI is the mean of the mean of the category calibration measures (**Table 2**; column 5) such that:

$$CI = \sum (CI_j / N)$$

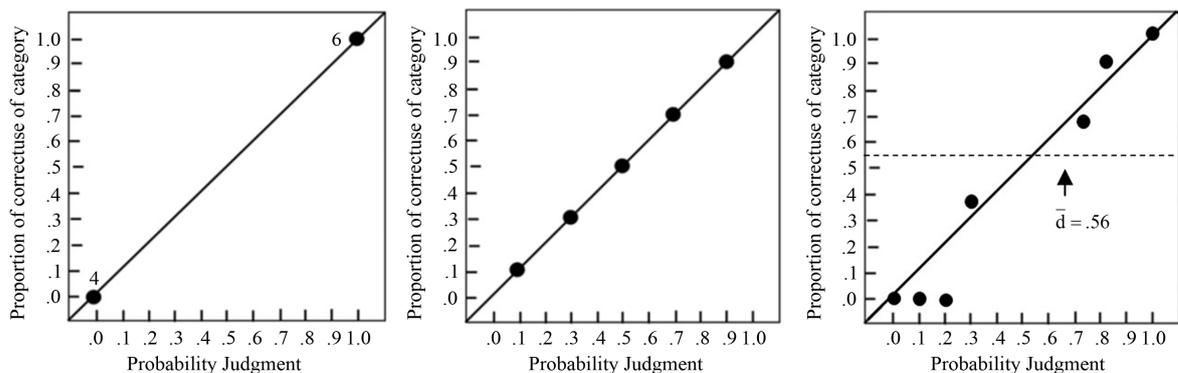


Figure 1. The left panel shows the case for a judgment with perfect calibration and discrimination to ten probability outcomes. The middle panel shows perfect calibration but not so good discrimination. The right panel shows an instance of a typical judge performance.

Table 1. Probability Score calculation to deception detection probability estimations.

Possible deceiver	Judge's Estimations (f)	Judge deception detection (f)			PS = (f - d) ²
		Lied?	Outcome Index value	f - d	
1	0.8	YES	1	-0.20	0.04
2	0.3	NO	0	0.30	0.09
3	0.7	YES	1	-0.30	0.09
4	0.9	YES	1	-0.10	0.01
5	0.8	NO	0	0.80	0.64
6	0.8	YES	1	-0.20	0.04
7	0.1	NO	0	0.10	0.01
8	0.3	YES	1	-0.70	0.49
9	0.7	NO	0	0.70	0.49
10	0.8	YES	1	-0.20	0.04
11	0.3	NO	0	0.30	0.09
12	0.0	NO	0	0.00	0.00
13	0.8	YES	1	-0.20	0.04
14	0.7	YES	1	-0.30	0.09
15	0.3	NO	0	0.30	0.09
16	0.9	YES	1	-0.10	0.01
17	0.7	YES	1	-0.30	0.09
18	0.7	NO	0	0.70	0.49
19	0.3	NO	0	0.30	0.09
20	0.8	YES	1	-0.20	0.04
21	0.2	NO	0	0.20	0.04
22	0.7	YES	1	-0.30	0.09
23	0.3	YES	1	-0.70	0.49
24	0.1	NO	0	0.10	0.01
25	0.7	YES	1	-0.30	0.09

On the other side a discrimination value is the mean of discrimination values for each category (Table 2; column 6):

$$DI = \sum \frac{DI_j}{N}$$

A software was implemented at our lab to compute these accuracy values for all study participants' probability estimations.

Computer videos of probably deceivers were implemented by using two short movies. First a movie containing positive contents and another with very negative content. Here, 16 volunteers were videotaped when they were lying or not lying about these movies. Each participant was appointed to a 30 minute session that consisted of five parts. First, debriefing was provided. Then they were sited in front of TV screen so they could watch both types of video (environmental relaxing images vs. stressful chirurgic contents). Each video lasted around 4 mi-

Table 2. An example of a procedure to obtain calibration and discrimination values.

(1)	(2)	(3)	(4)	(5)	(6)
Category (<i>j</i>)	f_j	N_j	\bar{d}_j	C_{ij}	D_{ij}
1	0.0	1	-	-	-
2	0.1	2	0.0	0.02	-
3	0.2	1	0.0	0.04	-
4	0.3	6	0.33	0.054	0.05
5	0.4	-	-	-	-
6	0.5	-	-	-	-
7	0.6	-	-	-	-
8	0.7	7	0.71	0.0007	0.0225
9	0.8	6	0.83	0.054	0.0729
10	0.9	2	1	0.02	0.1936
11	1.0	-	.	0.0	.
Mean				0.03145	0.13052
Exponent					
$CI_j = N_j (f_j - \bar{d}_j)^2 = (3)[(2) - (4)]^2$					
$DI_j = N_j (\bar{d}_j - \bar{d})^2 = (3)[(4) - (7)]^2$					
where (7) = $d = 0.56$					

minutes and were randomly presented. After each video presentation each participant was interviewed about the video contents while they were videotaped. The interviewer did not know what type of video they have just watched. Strict editing controls were carried over the videos of judges to avoid that environment distractors or video quality could interfere. **Figure 2** shows some examples of computer screen images from these videos.

In order to proceed with the experiment participants were sited in front of a computer that was connected to a server. The intention was to obtain a system that in the future can be used through internet. The software used for the study implementation through this LAN system was Inquisit v.3 by milliseconds software.

2.3. Procedure

As previously indicated each participant had to be seated in front of a computer. They were considered as judges whose experimental task was to decide if a person in the computer screen was telling the truth about a video content (e.g. if the video content was indeed positive or negative). Each judge had to observe 16 videos. These videos were randomly presented to each judge and each one lasted 4 minutes. Here, participants had to use the computer keyboard to provide a probability judgment between 0 and 1 to indicate if someone is lying. They had only 5 seconds to provide their answers once a video has finished.

3. Results

In order to proceed with the statistical analysis of the study's data a detection deception index error to each participant was obtained for their calibration and discrimination values. The idea was to obtain an estimate of how well participants' performance was in regard to both accuracy indexes. Since the CI values already relates to an error of judgment no correction was made. However, DI values were subtracted from 1 to indicate an error value. Then, a 3 (group: children, teens and adults) \times 2 (accuracy: calibration vs discrimination) mixed ANOVA was carried out over the participants' error values to each accuracy judgment.

Results showed a main effect for the accuracy factor $F(1, 24) = 8.637, p = 0.0071, \eta_p^2 = 0.2646$. However, no main effect was obtained for the age factor. If a deeper analysis is considered over the different age groups then a different judgment behavior is observed for each group. Here, analytical comparisons between calibration and discrimination behavior for each age group showed a significant difference between both indexes to children $F(1, 24) = 6.9023, p = 0.0147$ as well as for the teens group $F(1, 24) = 4.4812, p = 0.0448$. This was not the case for the adults group who showed no significant difference to this comparison $F(1, 24) = .2843, p = 0.5987$. **Figure 3** shows the interaction graph to both study factors.

No significant interaction between both factors was obtained $F(2, 24) = 0.98963, p = 0.38638$ but is interesting to notice that in contrast to judgment discrimination, judgment calibration error seems to increase in adulthood. This is supported by considering the global error index value PS. **Figure 4** Shows participants' performance under this scrutiny.

PS scores from **Figure 4** suggest that in the current study the children outperformed some adults and teenagers when it came the case to detect deception. Even when this is not always the case (supported by high calibration variability in **Figure 3**) it is hard to ignore this result. Increased sample size might bring insights over this result.

Regarding participants calibration graphs, **Figure 5** shows instances for the best and worst performance for each age group. Notice that is possible to immediately observe noticeable differences among participants. For instance notice the left top panel where a remarkable child had almost perfect calibration. On the other hand no-

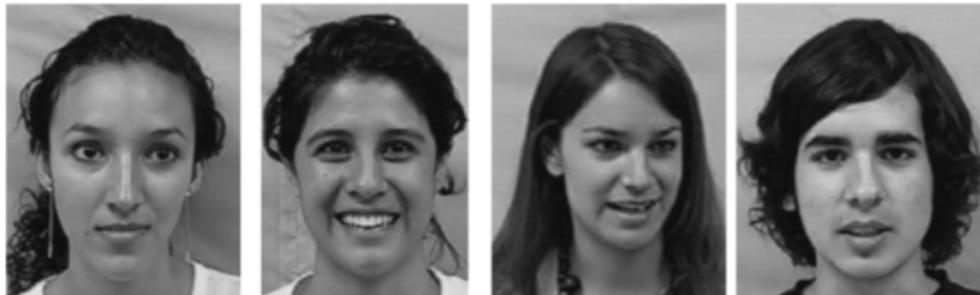


Figure 2. Snapshots of videos of possible deceivers describing the video content (positive or negative) they just watched.

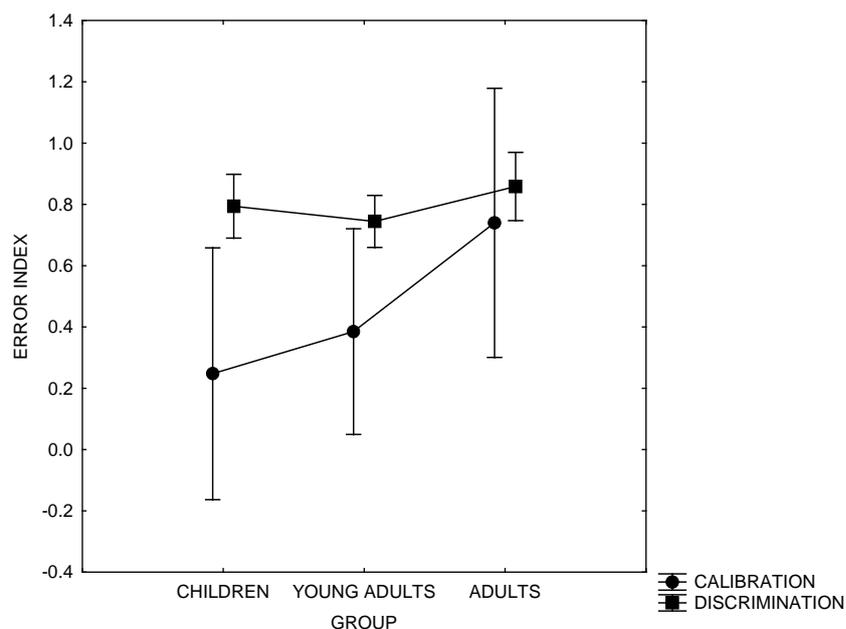


Figure 3. Interaction graph showing overall accuracy performance from the three age groups through all study experimental conditions.

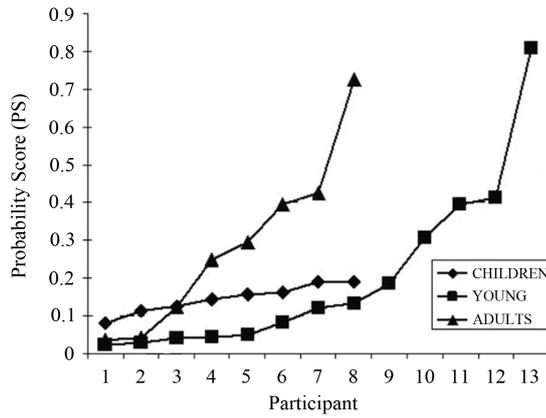


Figure 4. Global index error PS to deception detection for each participant in the study. Some adults obtained the highest index PS scores.

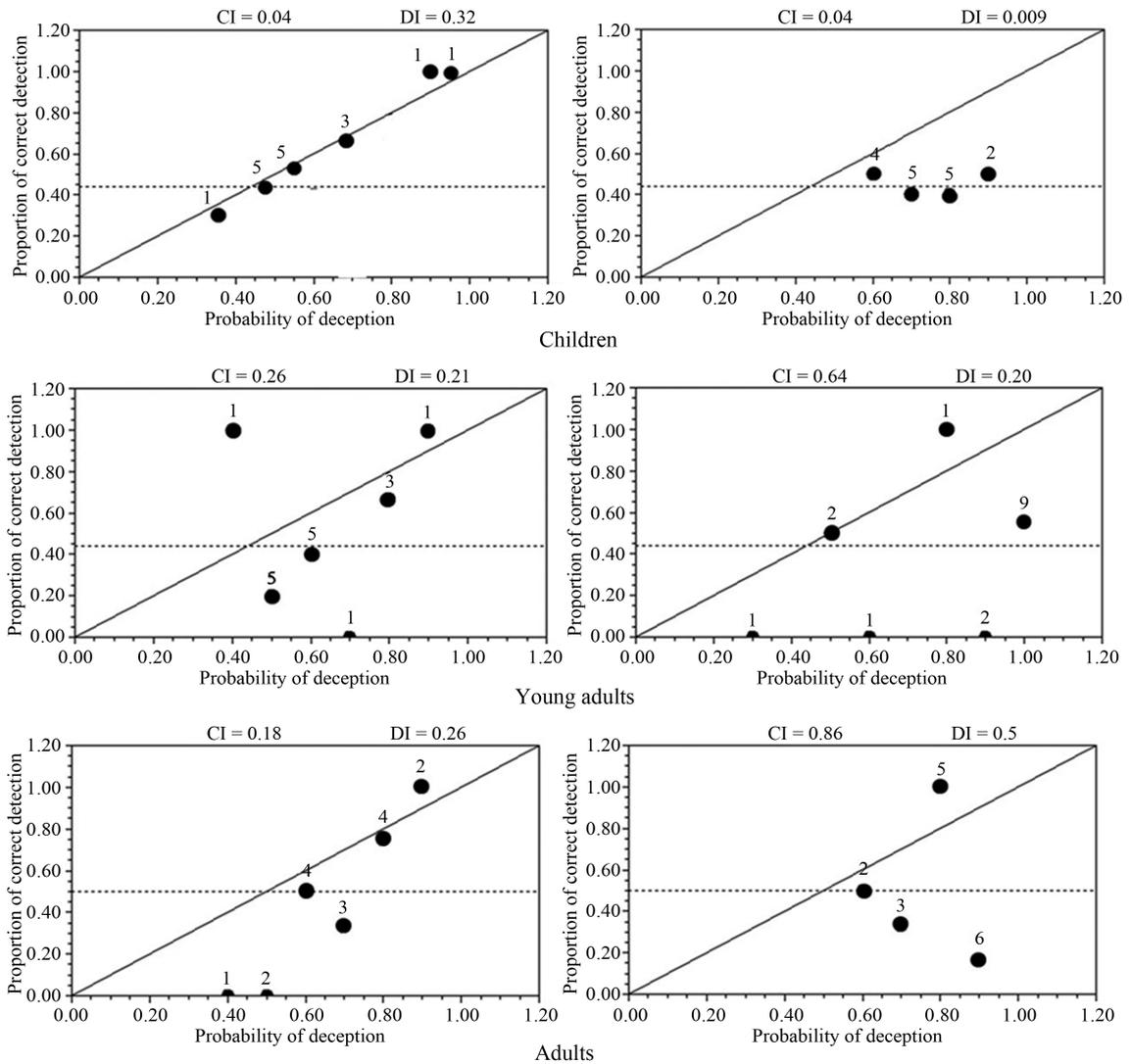


Figure 5. Examples of calibration graphs for each group. Only two instances of participation for each group (best (Left) and worst (Right) performance) are displayed. The number above each dot indicates the number of times a probability category was used.

tice how easy is to identify low accuracy (poor detection deception judgment) at the left panels from **Figure 5**. Overall, this study calibration graphs showed a variety of judgment styles to deception detection where each judgment style can be identified as belonging to a judge type reported in academic literature (e.g. Uniform or base rate judge type; e.g. Yates, 1990).

4. Discussion

It is clear from the current results that some children outperform some adults whenever a global accuracy index to deception detection is considered (PS). However, this might be related only to judgment calibration as it can be perceived in **Figure 3**. This means that in some instances of deception detection, children seem to be more properly qualified regarding to how sure they are to detect if someone is lying to them. This does not imply that children are always correct at discriminating a deceiving event in this study. Why that children sometimes do better at catching a liar is hard to know since most research efforts have addressed to explore age effects over the deceiver (e.g. Webb, 2006; Rachelle, Smith, & LaFreniere, 2013; Slessor, Phillips, Ruffman, Bailey, & Insch, 2014). As a matter of speculation it could be the case that children are not prone to be deceived by stereotypes that liars tend to use to achieve their deceiving goals. As it is suggested by the content analysis model to deception detection (Rogers, Boals & Drogin, 2011) or by the systematic thinking approach (Castro, Morales & Lopez, 2012), successful deceiving over adults results from activation of preconceived information in long term memory about possible lying behavior or biased cues of truthful behavior. Since children have not been sufficiently enough exposed to cultural deceiving behavior, then they do not react in the same way as adults.

However, what is clear from the current results analysis is the utility of not considering judgment accuracy to deception detection as a unitary cognitive concept. Rather, by using calibration graphs, judgment styles or judges types can be typified. Current deception detection methods are not capable of providing this cognitive information. For instance, the cognitive load model (Blandón-Gitlin, Pezdek, Lindsay, & Hagan, 2009) or the appointed content analysis model do relate to a structural cognitive approach rather than a functional consideration of judgment to deception detection. Further research is on demand to establish the complementary utility of this approach to current methods to catch a liar.

Finally, a limitation of the current study must be noticed since sample sizes are limited. Follow-up studies will explore this limitation. However, the main contribution of this study relies on having a new way to diagnose deception detection over three accuracy parameters, case by case using calibration graphs.

Acknowledgements

Financial support to this academic effort was provided from a Mexican government institution research grant named PROMEP.

References

- Blandón-Gitlin, I., Pezdek, K., Lindsay, S. D., & Hagan, L. (2009). Criteria-Based Content Analysis of True and Suggested Accounts of Events. *Applied Cognitive Psychology*, 23, 901-917. <http://dx.doi.org/10.1002/acp.1504>
- Bond Jr., C., & De Paulo, B. (2006). Accuracy of Deception Judgments. *Personality and Social Psychology Review*, 10, 214-234. <http://dx.doi.org/10.1086/286455>.
- Burgoon, J., Stoner, G., Bonito, J., & Dunbar, N. (2003). Trust and Deception in Mediated Communication. *Proceedings of the 36th Hawaii International Conference on Systems Sciences*, Maui, USA. <http://dx.doi.org/10.1109/HICSS.2003.1173792>
- Caso, L., Gnisci, A., Vrij, A., & Mann, S. (2005). Processes Underlying Deception: An Empirical Analysis of Truth and Lies When Manipulating the Stakes. *Journal of Investigative Psychology and Offender Profiling*, 2, 195-202. <http://dx.doi.org/10.1002/jip.32>
- Castro, C. Morales, G., & Lopez, E. (2012). Cognitive Algebra to Deception Detection: Information Integration Theory Contributions to Study Human Deception Detection's Cognitive Nature. *International Journal of Psychology Research*, Nova Science Publishers, Inc. 7. 1-11. ISSN: 1932-6092.
- DePaulo, B., Lindsay, J., Malone, B., Muhlenbruck, L., Charlton, K., & Cooper, H. (2003). Cues to Deception. *Psychological Bulletin*, 129, 74-112. <http://dx.doi.org/10.1037/0033-2909.129.1.74>
- Elfenbein, H., & Ambady, N. (2002). On the Universality and Cultural Specificity of Emotion Recognition: A Meta-Analysis. *Psychological Bulletin*, 128, 203-235. <http://dx.doi.org/10.1037/0033-2909.128.2.203>

- Elkins, D., Derrick, A., & Gariup, M. (2012). The Voice and Eye Gaze Behavior of an Imposter: Automated Interviewing and Detection for Rapid Screening at the Border. European Chapter of the Association for Computational Linguistics (EACL), Avignon, France.
- Farquhar, M. (2005). *A Treasury of Deception*. New York: Penguin Books.
- Fu, C. H., Williams, S. C., Brammer, M. J., Suckling, J., Kim, J., Cleare, A. J., Walsh, N. D., Mitterschiffthaler, M. T., Andrew, C. M., Pich, E. M., & Bullmore, E. T. (2007) Neural Responses to Happy Facial Expressions in Major Depression Following Antidepressant Treatment. *The American Journal of Psychiatry*, *164*, 599-602. <http://dx.doi.org/10.1176/appi.ajp.164.4.599>
- Gokhman, S., Hancock, J., Prabhu, P., Ott, M., & Cardie, C. (2012). In Search of a Gold Standard in Studies of Deception. *Proceedings of the EACL 2012 Workshop on Computational Approaches to Deception Detection*, Avignon, 23-27 April 2012.
- Hancock, J., Thom-Santelli, J., & Ritchie, T. (2004). Deception and Design: The Impact of Communication Technologies on Lying Behavior. *Proceedings of the Conference on Computer Human Interaction*, Vienna, 24-29 April 2004, 130-136. <http://dx.doi.org/10.1145/985692.985709>
- Hauch, V., Blandón-Gitlin, I., Massip, J., & Sporer, S. (2012). Linguistic Cues to Deception Assessed by Computer Programs: A Meta-Analysis. *Proceedings of the Workshop "Deception Detection" of the 13th Conference of the European Chapter of the Association for Computational Linguistics*, Avignon, 23-27 April 2012.
- Levine, T., Feeley, T., McCornack, S., Harms, C., & Hughes, M. (2005). Testing the Effects of Nonverbal Training on Deception Detection Accuracy with the Inclusion of a Bogus Train Control Group. *Western Journal of Communication*, *69*, Article ID: 203218.
- Rachelle, M., Smith, R., & LaFreniere, P. (2013). Development of Tactical Deception from 4 to 8 Years of Age. *British Journal of Developmental Psychology*, *31*, 30-41. <http://dx.doi.org/10.1111/j.2044-835X.2011.02071>
- Rogers, R., Boals, A., & Drogin, E. (2011). Applying Cognitive Models of Deception to National Security Investigations: Considerations of Psychological Research, Law, and Ethical Practice. *Journal of Psychiatry & Law*, *39*, 339-364.
- Rotenberg, K., & Sullivan, C. (2003). Children's Use of Gaze and Limb Movement Cues to Infer Deception. *Journal of Genetic Psychology*, *164*, 175-187. <http://dx.doi.org/10.1080/00221320309597976>
- Senter, S., & Dollins, A. (2003). *New Decision Rule Development: Exploration of a Two-Stage Approach*. Fort Jackson, SC: Department of Defense Polygraph Institute.
- Slessor, G., Phillips, L., Ruffman, T., Bailey, P., & Inch, P. (2014). Exploring Own-Age Biases in Deception Detection. *Cognition and Emotion*, *28*, 493-506. <http://dx.doi.org/10.1080/02699931.2013.839438>
- Sweeney, C., & Ceci, S. (2014). Deception Detection, Transmission, & Modality in Age & Sex. *Frontiers in Psychology*, *5*, 1-10. <http://dx.doi.org/10.3389/fpsyg.2014.00590>
- Trovillo, P. (1939) History of Lie Detection. *Journal of Criminal Law & Criminology*, *30*, 848-881.
- Webb, J. (2006). *Creating A New Reality: Information Age Effects on the Deception Process*. Ph.D. Thesis, Maxwell Air Force Base, AL: School of Advanced Air and Space Studies Air University.
- Yates, F. (1990). *Judgment and Decision Making*. Englewood Cliff, NJ: Prentice Hall.

Scientific Research Publishing (SCIRP) is one of the largest Open Access journal publishers. It is currently publishing more than 200 open access, online, peer-reviewed journals covering a wide range of academic disciplines. SCIRP serves the worldwide academic communities and contributes to the progress and application of science with its publication.

Other selected journals from SCIRP are listed as below. Submit your manuscript to us via either submit@scirp.org or [Online Submission Portal](#).

