Experience, Learning, and the Detection of Deception

Sanchaita Hazra[1]

with **Priyodorshi Banerjee**[2] *and* **Sanmitra Ghosh**[3]

University of Utah[1], Indian Statistical Institute Kolkata[2], and Jadavpur University[3]

Deception and Detection

- Suppose A can choose either X or Y
  X is beneficial to A and harmful to B, while Y is harmful to A and beneficial to B
  B knows which of X and Y is harmful to whom, while A does not

Deception and Detection

- ➔ Suppose A can choose either X or Y
  X is beneficial to A and harmful to B, while Y is harmful to A and beneficial to B
  B knows which of X and Y is harmful to whom, while A does not

- ➔ If B misreports information for influencing A to choose Y, then B is attempting to deceive A. B benefits from successful deception, A from successful detection of deception

Deception and Detection

- Suppose A can choose either X or Y
  X is beneficial to A and harmful to B, while Y is harmful to A and beneficial to B
  B knows which of X and Y is harmful to whom, while A does not

- If B misreports information for influencing A to choose Y, then B is attempting to deceive A. B benefits from successful deception, A from successful detection of deception

- Two questions arise:
  - Are there valid indicators of deception? (visual cues, change in tonality, speech/language)
  - Can deception be detected?

Deception and Detection

- ➔ Suppose A can choose either X or Y
  X is beneficial to A and harmful to B, while Y is harmful to A and beneficial to B
  B knows which of X and Y is harmful to whom, while A does not

- ➔ If B misreports information for influencing A to choose Y, then B is attempting to deceive A. B benefits from successful deception, A from successful detection of deception

- ➔ Two questions arise:
  - Are there valid indicators of deception? (visual cues, change in tonality, speech/language)
  - Can deception be detected?

- ➔ Raises current interest given increase in ‘fake news’, manipulated information, identity fraud, false testimony etc., which can have negative consequences
  (Rose 2017; Fujiwara et al 2021; Kim et al 2020; Institutional Investor 2019)

Deception and Detection: Theory and Practice

- Economic theory suggest if preferences of a sender and a receiver are opposed, strategic communication is uninformative, so deception in communication cannot be systematically detected (Crawford and Sobel 1982)

Deception and Detection: Theory and Practice

- ➔ Economic theory suggest if preferences of a sender and a receiver are opposed, strategic communication is uninformative, so deception in communication cannot be systematically detected (Crawford and Sobel 1982)

- ➔ In practice, senders may involuntarily generate behavioral cues helping to distinguish sincere and opportunistic communication
  (Zuckerman et al. 1981; DePaulo et al. 2003; George et al. 2004; Wang et al 2010; Hartwig and Bond 2011)

Deception and Detection: Theory and Practice

➔ Economic theory suggest if preferences of a sender and a receiver are opposed, strategic communication is uninformative, so deception in communication cannot be systematically detected (Crawford and Sobel 1982)

➔ In practice, senders may involuntarily generate behavioral cues helping to distinguish sincere and opportunistic communication
(Zuckerman et al. 1981; DePaulo et al. 2003; George et al. 2004; Wang et al 2010; Hartwig and Bond 2011)

➔ Receivers could benefit from being able to interpret these cues or valid indicators

Deception and Detection: Theory and Practice

- ➔ Economic theory suggest if preferences of a sender and a receiver are opposed, strategic communication is uninformative, so deception in communication cannot be systematically detected (Crawford and Sobel 1982)

- ➔ In practice, senders may involuntarily generate behavioral cues helping to distinguish sincere and opportunistic communication
  (Zuckerman et al. 1981; DePaulo et al. 2003; George et al. 2004; Wang et al 2010; Hartwig and Bond 2011)

- ➔ Receivers could benefit from being able to interpret these cues or valid indicators

- ➔ However, humans tend to be poor at deception detection, often performing at levels consistent with *random* decision making
  (Baesen et al 1948, Ockenfels and Selten 2002, Gneezy 2005, Serra-Garcia and Gneezy 2021)

Does Experience Matter?

- We investigate if experience as a factor can aid the detection of deception

- Extensive investigation of link between experience and learning
(Kraut and Poe 1980; Wang et al 2010)

  - Experience augments productivity (human capital models; Mincer 1974, Becker 1975)
  - Experience causes learning in labs (Newell and Rosenbloom, 1981, Erev and Haruvy, 2016)
  - For judges or law enforcement officers, length of active service determines seniority (DePaulo and Pfeifer 1986)

- In deception, individuals may be able to observe cues potentially associated with deception, with experience enabling learning of patterns, and improving inference

Our Paper

- ➔ We use game show data to study a high-stake, quasi-naturalistic, repeated decision-making environment with feedback

- ➔ We focus on situations where individuals may be repeatedly exposed to environments with potential deception, and ask if repeated exposure or experience can induce learning in presence of existing truthful information and thereby reduce error in detection

- ➔ **Contribution:** experience may be a valid factor aiding deception detection

The Game-show: To Tell The Truth

➔ We observe season 1 of a US tv game show To Tell the Truth dated Dec 1956 - Dec 1959

The Game-show: To Tell The Truth

- We observe season 1 of a US tv game show To Tell the Truth dated Dec 1956 - Dec 1959
- 145 episodes (averaging 25 minutes); each with multiple sessions, total of 429 sessions

The Game-show: To Tell The Truth

- We observe season 1 of a US tv game show To Tell the Truth dated Dec 1956 - Dec 1959

- 145 episodes (averaging 25 minutes); each with multiple sessions, total of 429 sessions

- A regular session: 1 host, 4 judges, 3 challengers.
  - Judges fixed for episode; can appear in multiple episodes
  - A challenger appears only once in the season

The Game-show: To Tell The Truth

- ➔ We observe season 1 of a US tv game show To Tell the Truth dated Dec 1956 - Dec 1959
- ➔ 145 episodes (averaging 25 minutes); each with multiple sessions, total of 429 sessions
- ➔ A regular session: 1 host, 4 judges, 3 challengers.
  - Judges fixed for episode; can appear in multiple episodes
  - A challenger appears only once in the season
- ➔ The 3 contestants comprise of 1 central character (CC) - about whom some absolute information is revealed (basis of detection), 2 imposters

The Game-show: To Tell The Truth

- ➔ We observe season 1 of a US tv game show To Tell the Truth dated Dec 1956 - Dec 1959

- ➔ 145 episodes (averaging 25 minutes); each with multiple sessions, total of 429 sessions

- ➔ A regular session: 1 host, 4 judges, 3 challengers.
  - ◆ Judges fixed for episode; can appear in multiple episodes
  - ◆ A challenger appears only once in the season

- ➔ The 3 contestants comprise of 1 central character (CC) - about whom some absolute information is revealed (basis of detection), 2 imposters

- ➔ The judges can publicly question the challengers
  The CC must speak the truth while the imposters impersonate CC and deceive the judges

The Game-show: To Tell The Truth

- We observe season 1 of a US tv game show To Tell the Truth dated Dec 1956 - Dec 1959

- 145 episodes (averaging 25 minutes); each with multiple sessions, total of 429 sessions

- A regular session: 1 host, 4 judges, 3 challengers.
  - Judges fixed for episode; can appear in multiple episodes
  - A challenger appears only once in the season

- The 3 contestants comprise of 1 central character (CC) - about whom some absolute information is revealed (basis of detection), 2 imposters

- The judges can publicly question the challengers
  The CC must speak the truth while the imposters impersonate CC and deceive the judges

- The task of each judge is to independently determine which of the challengers is the CC

Incentives for Judges and Challengers

- Incentive for the Judges (celebrity) to detect deception:
  - Viewership of the show comprised of actual or potential fans of the judges
    They should perform well in order to acquire, impress or retain followers

Incentives for Judges and Challengers

- Incentive for the Judges (celebrity) to detect deception:
  - Viewership of the show comprised of actual or potential fans of the judges
    They should perform well in order to acquire, impress or retain followers

- Incentive for the Challengers to deceive:
  - The challengers as a group got $250, to be divided equally,
    for every judge making a mistake in identification
  - A challenger could thus earn upto $333 if all incorrect votes were cast

Main Methodology

- There are 4 judgements per regular session
  - Some sessions (32) were not regular: irregularity is controlled for in the analysis (number of sessions/number of total challengers, different host, judge recused)

Main Methodology

- There are 4 judgements per regular session
  - Some sessions (32) were not regular: irregularity is controlled for in the analysis (number of sessions/number of total challengers, different host, judge recused)
- Outcome variable: **judge's decision**: either correct or incorrect; binary
- Independent variable: **number of sessions** (capturing the degree of experience) including the session from which the observation is being drawn

Main Methodology

- There are 4 judgements per regular session
  - Some sessions (32) were not regular: irregularity is controlled for in the analysis (number of sessions/number of total challengers, different host, judge recused)
- Outcome variable: **judge's decision**: either correct or incorrect; binary
- Independent variable: **number of sessions** (capturing the degree of experience) including the session from which the observation is being drawn
- We aim to identify and explain any possible correlation between these two variables

Descriptive Statistics

- 56 judges, 35 male, 21 female
- Appearance was unevenly distributed
- Half appeared in 5 sessions or less
- 9 appeared in more than 30 sessions
- Min number of sessions appeared in: 2
- Max number of sessions appeared in: 360

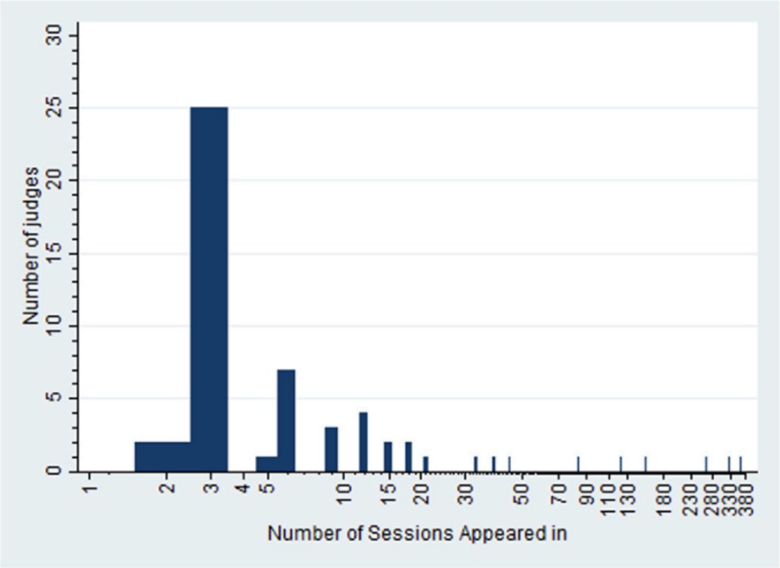

Fig. 1. Frequency distribution of appearances by judges.

Descriptive Statistics

- A Judge's error rate till (and including) the t^{th} appearance:

$$\left\{\frac{\text{No. of erroneous decisions till the } t^{th} \text{ appearance}}{\text{No. of decisions till the } t^{th} \text{ appearance}}\right\}$$

- The mean error rate across all judges in the full sample is 0.58, random decision error rate 0.67

- Mean error rate is lower than random benchmark level
  (two-sided p-values: Snedecor Cochran = 0.0007, t-test = 0.0703)

- **Error rate declines with experience**
  a (-)ve relationship bet. experience and error

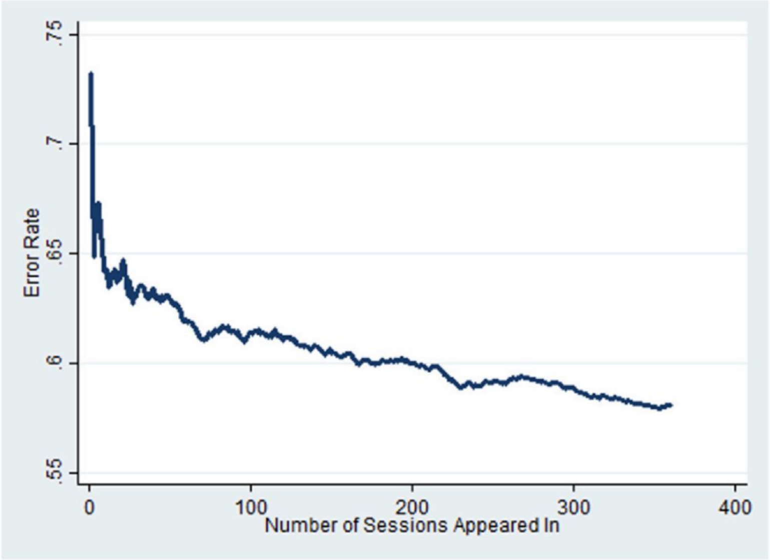

Fig. 2. Mean error rate.

Hypothesis

➔ **An immediate explanation is learning**
more experience allows better understanding and recognition of cues of deceptive behavior

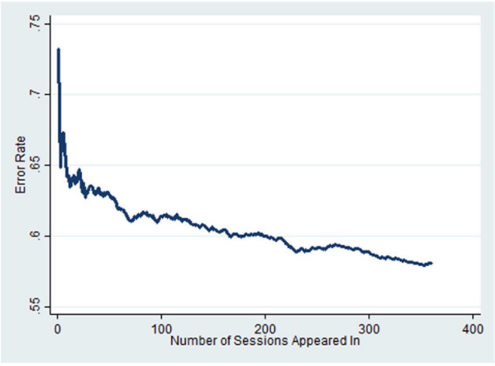

Fig. 2. Mean error rate.

Hypothesis

➔ **An immediate explanation is learning**
more experience allows better understanding and recognition of cues of deceptive behavior

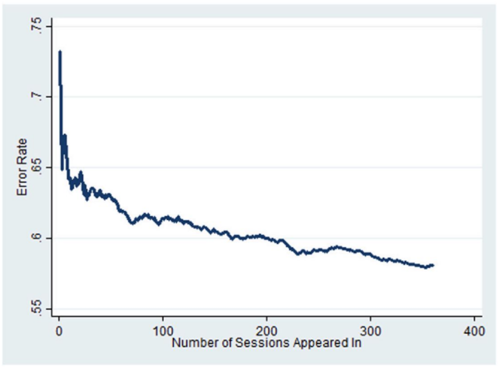

Fig. 2. Mean error rate.

➔ Two **alternatives** exist:
- **Alternative 1:** Maybe challenger groups got easier with time, i.e., later episodes (or sessions) involved innately easier challengers (whose deception was easier to detect)

Hypothesis

➔ **An immediate explanation is learning**
more experience allows better understanding and recognition of cues of deceptive behavior

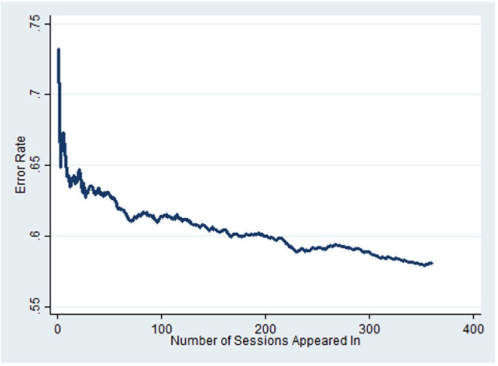

Fig. 2. Mean error rate.

➔ Two **alternatives** exist:

- **Alternative 1:** Maybe challenger groups got easier with time, i.e., later episodes (or sessions) involved innately easier challengers (whose deception was easier to detect)

- **Alternative 2:** Maybe the pool of judges selected in later episodes were innately better at detection. Alternative 2 is a particular concern as judges could be repeated, so selection of judges could be dependent on performance in prior episodes

Hypothesis

→ **An immediate explanation is learning**
more experience allows better understanding and recognition of cues of deceptive behavior

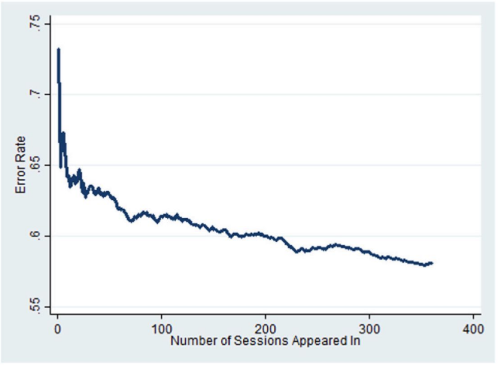

Fig. 2. Mean error rate.

→ Two **alternatives** exist:

- **Alternative 1:** Maybe challenger groups got easier with time, i.e., later episodes (or sessions) involved innately easier challengers (whose deception was easier to detect)

- **Alternative 2:** Maybe the pool of judges selected in later episodes were innately better at detection. Alternative 2 is a particular concern as judges could be repeated, so selection of judges could be dependent on performance in prior episodes

→ We focus on analyzing selection bias in judges and its correlation with the error trend with two analytical approaches

Analytical Approach I: Pooled Probit, Partial MLE

- ➔ Joint estimation of **performance** and **selection** of the judges controlling for unobserved characteristics of challenger groups (sessions/episodes)
- ➔ No distinction between self selection and producer selection for the judges
- ➔ No information on outside pools from which judges or challengers were drawn
- ➔ The model comprises of two equations:
  - ◆ Performance equation (PE), governing judges performance in any episode
  - ◆ Selection equation (SE), determining the selection of a judge in any episode

Table 1
Experience and learning: joint pooled probit estimates.

	Performance equation (1)		
	Dependent variable: judge decision (0 if correct)		
Focal regressor (t)	Appearance number	− 0.00249 * **	(0.00067)
Time-invariant judge-specific characteristics ($\mathbf{z}_i$)	Performance in first episode	0.48737 * *	(0.16098)
	Episode in which first appeared	− 0.00198	(0.00200)
	Age in days on first appearance	0.00004 * *	(0.00001)
	Gender (1 if male)	− 0.19964 * *	(0.07193)
Session level characteristics ($\mathbf{p}_{\sigma_{i,t}}$)	Number of female challengers	0.06833 * *	(0.02527)
	Total number of challengers	0.23496 * **	(0.05638)
	Recusal (1 if recusal)	− 0.41231	(0.38369)
Episode level characteristics ($\mathbf{w}_{e_{i,t}}$)	Host (1 if standard)	0.25269	(0.49477)
	Number of sessions in episode	− 0.24353	(0.57323)
Peer effect ($q_{\sigma_{i,t}}^{e_{i,t}}$)	Average peer cumulative performance	− 0.01240	(0.69188)
Session dummies ($\boldsymbol{\sigma}$)	Session 2	− 0.04068	(0.10770)
	Session 3	− 0.07095	(0.05223)
Episode dummies (**e**)	Yes		
	Selection equation (2)		
	Dependent variable: judge selection (1 if selected)		
Exclusion restrictions	Experience till previous episode ($x_{i,e}$)	0.00392 * **	(0.00033)
	Self cumulative peformance ($h_{i,e}$)	− 1.28870	(0.76851)
Time-invariant judge-specific characteristics ($\mathbf{z}_i$)	Performance in first episode	0.51720	(0.41517)
	Episode in which first appeared	− 0.00411	(0.00317)
	Age in days on first appearance	0.00004	(0.00004)
	Gender (1 if male)	− 0.33711	(0.34307)
Episode level characteristics ($\mathbf{w}_{e_{i,t}}$)	Host (1 if standard)	− 5.89751 * **	(1.63467)
	Number of sessions in episode	− 0.56489	(0.34895)
Peer effect ($q_{i,e}$)	Lagged average peer cumulative performance	9.24015 * **	(2.76535)
Episode dummies (**e**)	Yes		
Observations	9239		
Correlation coefficient (ρ)	− 0.10854		
p-value	0.06025		
Elasticity	− 0.19379		

Cluster adjusted standard errors in parentheses. $**p < 0.01$, $***p < 0.001$. For definitions of independent variables, see Appendix C.

Analytical Approach I: Results I

➔ Session and episode dummies are insignificant in the performance equation
No evidence for challengers differing in terms of ‘quality’ across sessions/episodes (Alt. 1)

Analytical Approach I: Results I

- ➔ Session and episode dummies are insignificant in the performance equation
  No evidence for challengers differing in terms of 'quality' across sessions/episodes (Alt. 1)

- ➔ (a) Intrinsic ability of judges influences overall error in the PE; (b) correlation between equations is weakly significant ($p = 0.6$) ➔ Possibility of performance dependent selection for judges; however, neither prior performance nor intrinsic ability are significant in the SE

- ➔ Correlation bet. intrinsic ability and no. of episodes of appearance was insignificant ($p = 0.8$)
  Overall, limited evidence in favor of sample selection effects

Analytical Approach I: Results I

➔ Session and episode dummies are insignificant in the performance equation
No evidence for challengers differing in terms of ‘quality’ across sessions/episodes (Alt. 1)

➔ (a) Intrinsic ability influences overall error in the PE; (b) correlation between equations is weakly significant ($p = 0.6$) ➔ Possibility of performance dependent selection for judges; however, neither prior performance nor intrinsic ability are significant in the SE

➔ Correlation bet. intrinsic ability and no. of episodes of appearance was insignificant ($p = 0.8$)
Overall, limited evidence in favor of sample selection effects

➔ Focal variable, ***Appearance number* is significant in the PE**
A doubling of experience leads to a 19% drop in the probability of error
Hence, **evidence in favor of learning**, after controlling for selection and arrangement effects

Analytical Approach I: Results II

- ➔ Higher error may be produced by higher age ceteris paribus
- ➔ Male judges had lower error in detection
- ➔ Female challengers were more successful at deception

Time-invariant	Performance in first episode	0.48737 * *	(0.16098)
judge-specific	Episode in which first appeared	− 0.00198	(0.00200)
characteristics	Age in days on first appearance	0.00004 * *	(0.00001)
($\mathbf{z}_i$)	Gender (1 if male)	− 0.19964 * *	(0.07193)
Session level	Number of female challengers	0.06833 * *	(0.02527)

Analytical Approach I: Robustness and Limitations

Robustness

- → No evidence that later episodes or later sessions within episodes were easier
- → Mixed evidence for performance dependent selection
- → Strong evidence in favor of learning using analytic standard errors
  Weak evidence using bootstrapped standard errors (p = 7%)

Analytical Approach I: Robustness and Limitations

Robustness

- ➔ No evidence that later episodes or later sessions within episodes were easier
- ➔ Mixed evidence for performance dependent selection
- ➔ Strong evidence in favor of learning using analytic standard errors
  Weak evidence using bootstrapped standard errors (p = 7%)

Limitations

- ➔ Alt. 2 not fully refuted: some evidence for selection bias, some evidence for learning
- ➔ No information on outside pool from which sample of 56 judges could have been drawn
- ➔ No information on possible dynamic internal pool from which judges were selected
- ➔ Methodological limitations regarding peer group selection of the judges

Analytical Approach II: Intra-Episode Analysis

- ➔ Judges were fixed for an episode, so intra-episode analysis eliminates effects of selection bias
- ➔ Focal variable (experience) becomes collinear with session dummies, so effects of unobserved session-level characteristics cannot be controlled for: but no effect in Analytical Approach I

Analytical Approach II: Intra-Episode Analysis

- ➔ Judges were fixed for an episode, so intra-episode analysis eliminates effects of selection bias

- ➔ Focal variable (experience) becomes collinear with session dummies, so effects of unobserved session-level characteristics cannot be controlled for: but no effect in Analytical Approach I

- ➔ 2 phases of Analytical Approach II:
  - ◆ **A. Learning over the episode of first appearance**
    - First appearance episode for each judge: no performance history

Analytical Approach II: Intra-Episode Analysis

- Judges were fixed for an episode, so intra-episode analysis eliminates effects of selection bias

- Focal variable (experience) becomes collinear with session dummies, so effects of unobserved session-level characteristics cannot be controlled for: but no effect in Analytical Approach I

- 2 phases of Analytical Approach II:
  - **A. Learning over the episode of first appearance**
    - First appearance episode for each judge: no performance history
  - **B. Average Intra-Episode Learning**
    - We construct an average episode:
      For each judge, we consider the average outcome over all first sessions in the episodes of appearance, all 2nd sessions in the episodes of appearance, and so on, to check if learning occurs through the course of an average episode

Analytical Approach II(A): Learning over the episode of first appearance

- ➔ Only from first appearance episode for each judge: no performance history
- ➔ Focal regressor is strongly significant with episode index variable, weakly with episode dummy variables (p = 0.7%)
- ➔ A doubling of experience leads to a 15% drop in the probability of error — **Favors learning**

Table 2: Intra-episode learning: first episode of appearance

		Dependent variable: judge decision (0 if correct)		
		pooled probit (episode index)	pooled probit (episode dummies)	correlated random effects
Focal	Session	-0.26611*	-0.35760	-0.25996*
regressor (t)		(0.12514)	(0.20001)	(0.12495)
Time	Episode in which	0.00113	-	0.00265
invariant	first appeared	(0.00342)	(-)	(0.00364)
judge	Age in days on	0.00001	-0.00040***	0.00002
specific	first appearance	(0.00004)	(0.00010)	(0.00004)
characteristics	Gender	-0.41848	-0.76087	-0.47715
$(\mathbf{z}_i)$	(1 if male)	(0.25844)	(0.51680)	(0.27333)
Session level	Number of	-0.04232	-0.09556	-0.06342
	female challengers	(0.08079)	(0.13498)	(0.08381)
characteristics	Total number	1.00399***	1.26159***	1.07998***
	of challengers	(0.13242)	(0.19282)	(0.25059)
$(\mathbf{p}_{\sigma_{i,t}})$	Recusal	-1.17168	-33.59296	-0.84677
	(1 if recusal)	(0.71291)	(.)	(0.44135)
Episode	Host	0.49018	-38.67860***	-0.11284
level	(1 if standard)	(0.39805)	(0.87498)	(0.43307)
characteristics	Number of sessions	0.48347	0.81918	0.07839
$(\mathbf{w}_{e_{i,t}})$	in episode	(0.28279)	(1.03288)	(0.74570)
Peer effect	Average peer	-0.91036	-3.72099	3.80848
$(q_{i,t})$	cum. perf.	(1.73334)	(4.33897)	(4.26399)
Judge averages		-	-	Yes
Episode dummies		-	Yes	-
Observations†		156	156	156
(Semi) Elasticity		-0.15356	0	-0.15017

†Values for the peer effect variable are not defined for the first session of the first episode of the show. We hence lose 4 observations from 160, leaving us with a total of 156.

Cluster adjusted standard errors in parentheses. * $p < 0.05$, *** $p < 0.001$.

Analytical Approach II(B): Average Intra-Episode Learning

- ➔ Focal regressor is marginally insignificant (p = 0.051)
- ➔ Probability of error reduces by 0.06 on average from one session to the next — **Favors learning**

Table 3: Intra-episode learning: all episodes

		Dependent variable: average judge decision
Focal regressor (t)	Session	-0.06676
		(0.03417)
Time-invariant	Performance in	0.67211***
	first episode	(0.07624)
judge-specific	Episode in which	-0.00012
	first appeared	(0.00045)
characteristics	Age in days on	0.00001
	first appearance	(0.00001)
$(\mathbf{z}_i)$	Gender	-0.02073
	(1 if male)	(0.04144)
Observations		165

Cluster adjusted standard errors in parentheses. *** $p < 0.001$.

Conclusion

- Bleak existing evidence suggesting that humans are not necessarily highly skilled at deception detection; leading to search for valid indicators of deception which may improve detection

- Our evidence suggests the **presence of learning via experience can improve performance**

- This may imply:
  - less experienced enforcement officers could be less successful on average at detecting criminal deception ceteris paribus
  - individuals may be more susceptible to detecting misinformation (lies/fake news) if they have lower prior exposure to deception environments
  - Inexperience may render a person more vulnerable if targeted for impersonation fraud (romance, finance)

Published, July, 2023

Journal of Economic Criminology 1 (2023) 100010

Contents lists available at ScienceDirect

Journal of Economic Criminology

journal homepage: www.journals.elsevier.com/journal-of-economic-criminology

Experience, learning and the detection of deception☆

Priyodorshi Banerjee[a,*], Sanmitra Ghosh[b], Sanchaita Hazra[c]

[a] *Economic Research Unit, Indian Statistical Institute, Kolkata, West Bengal, India*
[b] *Department of Economics, Jadavpur University, Kolkata, West Bengal, India*
[c] *Department of Economics, University of Utah, Salt Lake City, UT, USA*

Future Prospects

- ➔ A question that remains unanswered in our investigation is *what it is that is learnt*
- ➔ Additionally, each judge, while an independent decision maker, was always in the company of other judges in the show: more specific setups can be explored

Co-authors

Priyadarshi Banerjee

Sanmitra Ghosh

Thanks!

sanchaitahazra.com

@hsanchaita