Combining individuating and context-general cues in lie detection

David Peebles

July 26, 2024

University of Huddersfield

Centre for Cognition and Neuroscience

Talk outline

- The Adaptive Lie Detector theory (ALIED: Street, 2015)
- The ACT-R cognitive architecture (Anderson, 2007)
- Grounding ALIED in the representations and mechanisms of ACT-R

The Adaptive Lie Detector (ALIED) theory

ALIED: Main assumptions

- Judgements informed by two types of information:
  - Individuating (II). Cues related to particular statement under consideration
  - Context-general (CGI). Applies across statements and contexts. Subjective honesty base rate in current context

ALIED: Main assumptions

- Judgements informed by two types of information:
  - Individuating (II). Cues related to particular statement under consideration
  - Context-general (CGI). Applies across statements and contexts. Subjective honesty base rate in current context
- II and CGI weighted based on perceived diagnosticity

ALIED: Main assumptions

- Judgements informed by two types of information:
  - Individuating (II). Cues related to particular statement under consideration
  - Context-general (CGI). Applies across statements and contexts. Subjective honesty base rate in current context
- II and CGI weighted based on perceived diagnosticity
- Diagnosticity of II varies:
  - High (e.g., Pinocchio's nose grows) $\longrightarrow$ weight II more for high accuracy (Blair et al., 2010; Levine & McCornack, 2014)
  - Low (e.g., poker face) $\longrightarrow$ weight prior CGI ("most people tell the truth in this setting") more

ALIED's account of "truth bias"

- People tend to believe information to be true (C. F. Bond & DePaulo, 2006; Levine, 2014)

ALIED's account of "truth bias"

- People tend to believe information to be true (C. F. Bond & DePaulo, 2006; Levine, 2014)
- ALIED – typical situations:
  - Individuating cues typically have low diagnosticity
  - CGI people are generally truthful (Halevy et al., 2014)
  - Therefore, rational in most situations to assume truth

ALIED's account of "truth bias"

- People tend to believe information to be true (C. F. Bond & DePaulo, 2006; Levine, 2014)
- ALIED – typical situations:
  - Individuating cues typically have low diagnosticity
  - CGI people are generally truthful (Halevy et al., 2014)
  - Therefore, rational in most situations to assume truth
- ALIED – atypical situations:
  - Where lying (or belief that lying) is more prevalent
  - Bias is to assume that statements more likely to be false (G. D. Bond et al., 2005; Masip et al., 2009)

ALIED's account of "truth bias"

- People tend to believe information to be true (C. F. Bond & DePaulo, 2006; Levine, 2014)
- ALIED – typical situations:
  - Individuating cues typically have low diagnosticity
  - CGI people are generally truthful (Halevy et al., 2014)
  - Therefore, rational in most situations to assume truth
- ALIED – atypical situations:
  - Where lying (or belief that lying) is more prevalent
  - Bias is to assume that statements more likely to be false (G. D. Bond et al., 2005; Masip et al., 2009)
- Truth bias not a cognitive disposition but an adaptive judgement in absence of diagnostic individuating cues

Empirically testing the ALIED theory

- Street et al. (2016) investigated interaction between individuating and context-general information
- Ps given game-playing scenario where people could cheat and then be truthful or lie when later questioned
- Three components:

Empirically testing the ALIED theory

- Street et al. (2016) investigated interaction between individuating and context-general information
- Ps given game-playing scenario where people could cheat and then be truthful or lie when later questioned
- Three components:

- Training. Ps learn to associate four behavioural cues with probability of lying/telling truth (between 20% and 80%)
  - Voice pitch
  - Facial expression
  - Number of silent periods in sentences
  - Number of self-references such as 'I' and 'me'

Empirically testing the ALIED theory

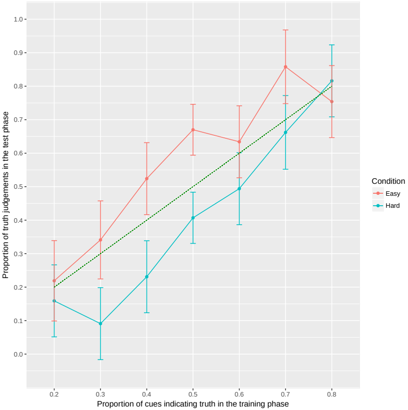
- Street et al. (2016) investigated interaction between individuating and context-general information
- Ps given game-playing scenario where people could cheat and then be truthful or lie when later questioned
- Three components:
- Suggest truth/lie base-rates. Ps told game was:
  - Easy (i.e., less cheating/lying)
  - Hard (i.e., more cheating/lying)

Empirically testing the ALIED theory

- Street et al. (2016) investigated interaction between individuating and context-general information
- Ps given game-playing scenario where people could cheat and then be truthful or lie when later questioned
- Three components:

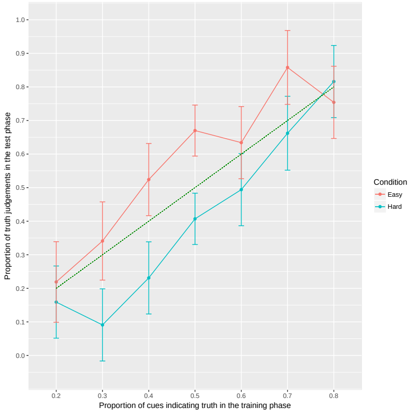
- Test. Ps presented with cues again and required to respond whether they indicated truth or lie

ALIED's predictions supported



Proportion of truth judgements for each cue diagnosticity in the test phase

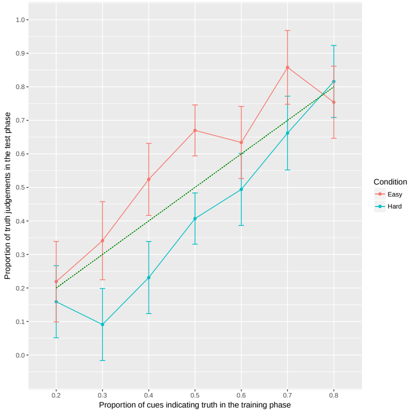
ALIED's predictions supported



Proportion of truth judgements for each cue diagnosticity in the test phase

- Truth judgements increase as cues are more indicative of honesty

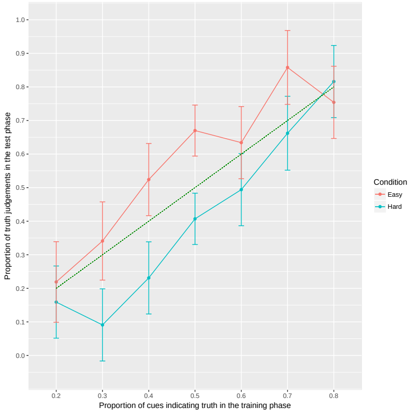
ALIED's predictions supported



Proportion of truth judgements for each cue diagnosticity in the test phase

- Truth judgements increase as cues are more indicative of honesty
- Context information shifts judgements in predicted directions

ALIED's predictions supported



Proportion of truth judgements for each cue diagnosticity in the test phase

- Truth judgements increase as cues are more indicative of honesty
- Context information shifts judgements in predicted directions
- Effect of CGI increases as the individuating cue diagnosticity decreases

Developing a mechanistic account

- Demonstrates how judgements arise from interaction of:
  - Information about the diagnosticity of individuating cues
  - Context-general information about the prevalence of lying

Developing a mechanistic account

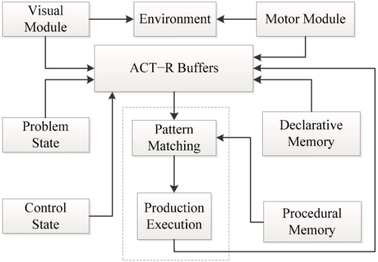
- Demonstrates how judgements arise from interaction of:
  - Information about the diagnosticity of individuating cues
  - Context-general information about the prevalence of lying
- Questions
  - How are the two types of information learned and cognitively represented?
  - What cognitive mechanisms can account for interaction?

Developing a mechanistic account

- Demonstrates how judgements arise from interaction of:
  - Information about the diagnosticity of individuating cues
  - Context-general information about the prevalence of lying
- Questions
  - How are the two types of information learned and cognitively represented?
  - What cognitive mechanisms can account for interaction?
- Cognitive process model
  - Developed within the ACT-R theory (Anderson, 2007)
  - Explains performance in terms of basic learning and retrieval mechanisms of declarative memory
  - Provides algorithmic level account consistent with ALIED

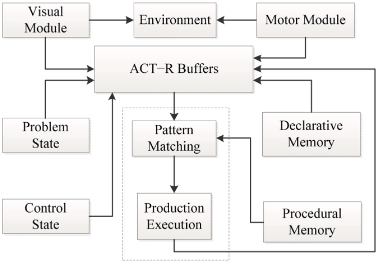
The ACT-R cognitive architecture

Key components of ACT-R



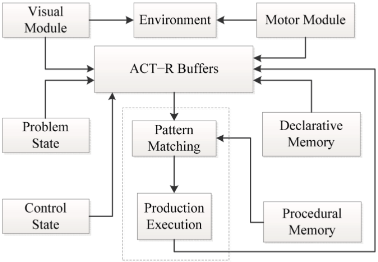
- *Core:* Two computational representations of memory
  - *Declarative* Network of "chunks" representing facts
  - *Procedural* "Production rules" representing actions
- Modules to simulate vision, audition, and motor action to interact with task environments

Key components of ACT-R



- *Rule-based sequential behaviour*
    - Every 50ms, snapshot of all buffer contents (goal state, visual object, retrieved knowledge etc.) is taken
    - Production rules matching buffer contents compete to "fire". Winner executes its actions (e.g., memory retrieval, motor actions, eye movements, update goal)

Key components of ACT-R



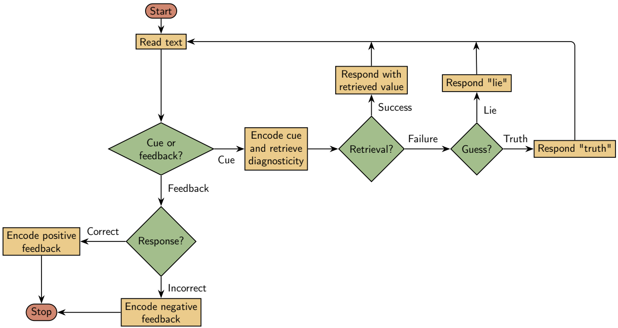
- *Equations that govern learning and forgetting*
  - Production rule "utility" learning. Productions involved in successful actions are reinforced
  - Chunk "activation" determines probability and speed of retrieval, forgetting etc.

Retrieving knowledge chunks from declarative memory

$$A_i = B_i + \sum_{j \in C} W_j S_{ji} + \sum_{l} PM_{li} + \epsilon$$

- Base-level activation reflects recency and frequency
  - Most recently and frequently used chunks have higher activation
- Partial matching component from retrieval cue
  - Retrievals don't require a perfect match to the cue
  - Chunks given a mismatch penalty based on similarity
- Noise component increases likelihood of erroneous response of chunk unrelated to retrieval cues

An ACT-R model of the experiment



- Model interacts with simulation of the experiment
- Code: github.com/djpeebles/act-r-lie-detection-model

An ACT-R model of the experiment

Before training

- 4 behavioural cues, differently diagnostic of truth/lie
- 8 chunks in declarative memory
- 2 per cue – one associated with “lie”, the other “truth”

Chunk	Activation
(voice-pitch truth)	0.0
(voice-pitch lie)	0.0
(facial-expression truth)	0.0
(facial-expression lie)	0.0
(silent-periods truth)	0.0
(silent-periods lie)	0.0
(self-references truth)	0.0
(self-references lie)	0.0

An ACT-R model of the experiment

During training

- Learn to associate cues with "true" and "lie" responses
- Use cue to retrieve associated chunks and make response
- Adjust chunk activations based on feedback

Chunk	Activation
(voice-pitch truth)	0.2
(voice-pitch lie)	0.0
(facial-expression truth)	0.1
(facial-expression lie)	0.3
(silent-periods truth)	0.4
(silent-periods lie)	0.0
(self-references truth)	0.1
(self-references lie)	0.0

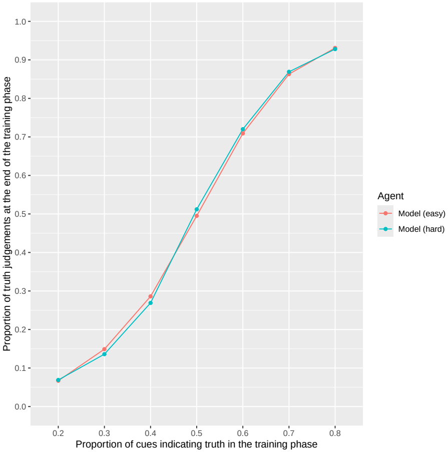
An ACT-R model of the experiment

After training

- Chunk activations reflect learned associations between cues and responses
- Cue diagnosticity
  - High - large difference between true/lie chunks
  - Low - small difference between true/lie chunks

Chunk	Activation
(voice-pitch truth)	0.8
(voice-pitch lie)	0.2
(facial-expression truth)	0.3
(facial-expression lie)	0.7
(silent-periods truth)	0.4
(silent-periods lie)	0.6
(self-references truth)	0.5
(self-references lie)	0.5

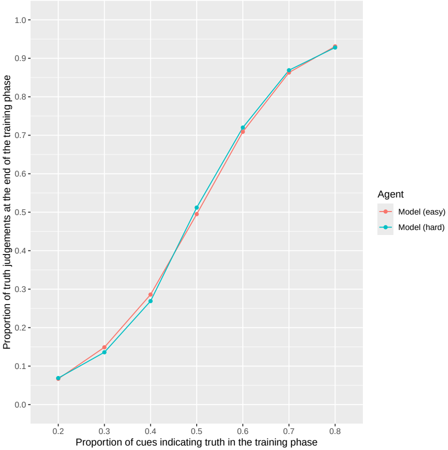
ACT-R performance after the training phase



Proportion of truth judgements for each cue diagnosticity after the training phase

- Model over- and under-estimates truthful statement proportions as cue diagnosticity increases
- Due to non-linearities in ACT-R's equations, differences in activation between competing chunks

ACT-R performance after the training phase



Proportion of truth judgements for each cue diagnosticity after the training phase

- Consistent with human probability learning with feed-back.
- People maximise responses rather than probability match (e.g., Barron & Erev, 2003; Shanks et al., 2002)

Providing the context information

- Between training and test, model provided condition information, “easy” or “hard”

Chunk	Activation
(voice-pitch truth)	0.8
(voice-pitch lie)	0.2
(facial-expression truth)	0.3
(facial-expression lie)	0.7
(silent-periods truth)	0.4
(silent-periods lie)	0.6
(self-references truth)	0.5
(self-references lie)	0.5

Providing the context information

- Between training and test, model provided condition information, “easy” or “hard”
- Model retrieves from memory associated context-general response bias (“truth” or “lie” respectively)

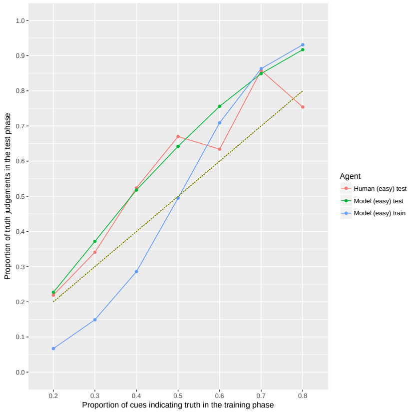
Chunk	Activation
(voice-pitch truth)	0.8
(voice-pitch lie)	0.2
(facial-expression truth)	0.3
(facial-expression lie)	0.7
(silent-periods truth)	0.4
(silent-periods lie)	0.6
(self-references truth)	0.5
(self-references lie)	0.5

Providing the context information

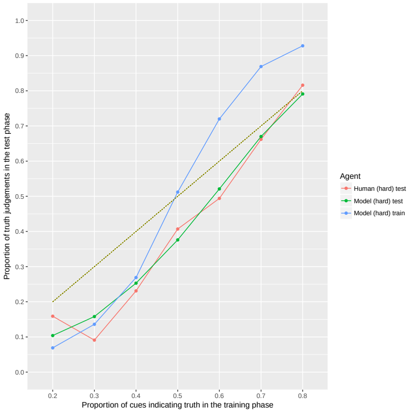
- Between training and test, model provided condition information, “easy” or “hard”
- Model retrieves from memory associated context-general response bias (“truth” or “lie” respectively)
- Response bias becomes an additional cue for retrievals in test phase

Chunk	Activation
(voice-pitch truth)	0.8
(voice-pitch lie)	0.2
(facial-expression truth)	0.3
(facial-expression lie)	0.7
(silent-periods truth)	0.4
(silent-periods lie)	0.6
(self-references truth)	0.5
(self-references lie)	0.5

Comparing human and model performance



"Easy" condition. $R^2 = 0.92, RMSD = 0.08$



"Hard" condition. $R^2 = 0.98, RMSD = 0.04$

Conclusions

- The ACT-R model is a process-level account of the human data consistent with ALIED theory

Conclusions

- The ACT-R model is a process-level account of the human data consistent with ALIED theory
- Demonstrates how learned diagnostic cues interact with context-general information

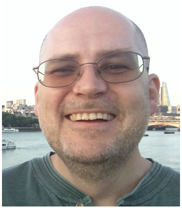
Conclusions

- The ACT-R model is a process-level account of the human data consistent with ALIED theory
- Demonstrates how learned diagnostic cues interact with context-general information
- Effect of CGI related to strength of diagnosticity
  - CGI has greater effect as diagnosticity of individuating cue reduces
  - CGI has weaker effect with strongly diagnostic cues

Conclusions

- The ACT-R model is a process-level account of the human data consistent with ALIED theory
- Demonstrates how learned diagnostic cues interact with context-general information
- Effect of CGI related to strength of diagnosticity
  - CGI has greater effect as diagnosticity of individuating cue reduces
  - CGI has weaker effect with strongly diagnostic cues
- Model supports compensatory strategy of integrating multiple cues rather than using only one (Gigerenzer & Todd, 1999; Newell & Shanks, 2003)

Acknowledgements



Chris Street, Keele University, UK

Dan Bothell, CMU, USA

References i

Anderson, J. R. (2007). *How can the human mind occur in the physical universe?* Oxford University Press.

Barron, G., & Erev, I. (2003).*Small feedback-based decisions and their limited correspondence to description-based decisions.* *Journal of Behavioral Decision Making*, *16*(3), 215–233.

Blair, J. P., Levine, T. R., & Shaw, A. S. (2010).*Content in context improves deception detection accuracy.* *Human Communication Research*, *36*(3), 423–442.

Bond, C. F., & DePaulo, B. M. (2006).*Accuracy of deception judgments.* *Personality and social psychology Review*, *10*(3), 214–234.

Bond, G. D., Malloy, D. M., Arias, E. A., Nunn, S. N., & Thompson, L. A. (2005).*Lie-biased decision making in prison.* *Communication Reports*, *18*(1–2), 9–19.

Gigerenzer, G., & Todd, P. M. (1999). *Simple heuristics that make us smart.* Oxford University Press, USA.

Halevy, R., Shalvi, S., & Verschuere, B. (2014).*Being honest about dishonesty: Correlating self-reports and actual lying.* *Human Communication Research*, *40*(1), 54–72.

Levine, T. R. (2014).*Truth-Default Theory (TDT): A theory of human deception and deception detection.* *Journal of Language and Social Psychology*, *33*(4), 378–392.

Levine, T. R., & McCornack, S. A. (2014).*Theorizing about deception.* *Journal of Language and Social Psychology*, *33*(4), 431–440.

Masip, J., Alonso, H., Garrido, E., & Herrero, C. (2009).*Training to detect what? The biasing effects of training on veracity judgments.* *Applied Cognitive Psychology*, *23*(9), 1282–1296.

Newell, B. R., & Shanks, D. R. (2003).*Take the best or look at the rest? Factors influencing "one-reason" decision making.* *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*(1), 53–65.

Shanks, D. R., Tunney, R. J., & McCarthy, J. D. (2002).*A re-examination of probability matching and rational choice.* *Journal of Behavioral Decision Making*, *15*(3), 233–250.

Street, C. N. H. (2015).*ALIED: Humans as adaptive lie detectors.* *Journal of Applied Research in Memory and Cognition*, *4*(4), 335–343.

Street, C. N. H., Bischof, W. F., Vadillo, M. A., & Kingstone, A. (2016).*Inferring others' hidden thoughts: Smart guesses in a low diagnostic world.* *Journal of Behavioral Decision Making*, *29*(5), 539–549.