Department of Psychiatry
SCHOOL OF MEDICINE
UNIVERSITY OF COLORADO **ANSCHUTZ MEDICAL CAMPUS**

# Learning Types during Interpretation Bias Training to Treat Irritability

Tara Trujillo[1], Joel Stoddard[1], Simone Haller[2], Melissa Brotman[2], Matt Jones[3]

[1] Children's Hospital Colorado, Anschutz Medical Campus
[2]National Institutes of Health
[3] Department of Psychology and Neuroscience, University of Colorado Boulder

Children's Hospital Colorado

## Background

Chronic irritability is a common and impairing complaint in pediatric psychopathology for which evidence-based treatments are scarce, especially those that target the neurocognitive basis of irritability.

Some psychopathologies are associated with biases when judging ambiguous social stimuli. A potential treatment target for chronic irritability is hostile interpretation bias (HIB), a tendency to interpret ambiguous stimuli as threatening. HIB may be targeted and changed via a computer-based interpretation bias training (IBT).

This treatment has been controversial because the mechanism of action has not been well studied. However, IBT is amenable to the application of category learning theory to measure behavioral and neural changes. Given variable success of IBT in prior trials, we hypothesized that those in active learning conditions in IBT may be learning via distinctive mechanisms.

We expect that phenomena associated with anxiety and irritability will influence category learning during IBT, especially:

1) A bias towards judging ambiguous stimuli as hostile.
2) An attentional bias towards threatening stimuli.
3) Impairments in labeling face emotions.

## Methods

A session of IBT learning was assessed in 63 transdiagnostic youth with varying severity of irritability and anxiety who would meet criteria for being in a single active training group in clinical trials of IBT.

### IBT Design

Participants judge a continuum of facial expressions, ranging on 2 dimensions: neutral to happy and neutral to angry. Then, they train in IBT towards less angry and more happy judgments of ambiguous faces on this continuum.

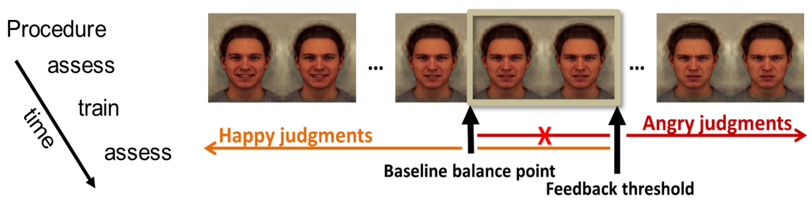


***Figure 1.*** An interpretation bias training session. First the balance point is assessed, which is the point on a continuum of happy to angry faces where a person's judgments of them switch from predominantly happy to angry. Training sessions provide feedback on judgments that encourage happy judgments of two ambiguous faces previously judged as angry. Training effects are measured by another balance point assessment.

### Computational Model

ALCOVE is a model of category learning that accounts for similarity between stimuli along multidimensional continua. It was applied to each person's training data. Three major model-based parameters were of interest:

| | |
|---|---|
| Indifference Point | The point along the morph continuum where the probability of an angry judgement is 0.5. Relatively lower values indicate HIB. |
| Generalization | The degree to which a stimulus activates representations of neighboring morphs. Larger values reflect less specificity when applying learned updates. |
| Effective Learning Rate | Learning rate adjusted by the degree of generalization to facilitate comparisons between individuals. |

We assessed associations with psychopathology (anxiety and irritability), dimensionally via multivariate linear models. We also empirically assessed for types of learners using generalized mixture models.

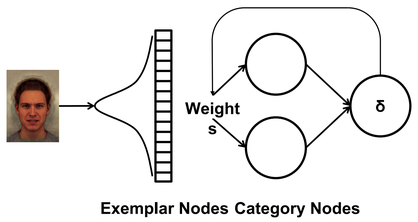


***Fig 2***. Schematic of ALCOVE, generalization, represented by the bell curve, and error driven learning represented by the adaptive weights. ALCOVE required adaptations for IBT and patients which were validated in an independent sample.

## Results

The computational model to assess learning during IBT demonstrates different types of IBT learners with unique associations to age and psychopathology.

In multivariate linear modeling, individuals with higher generalization tended towards anxiety (b=1.7 (0.9), p=.05) and were younger (b=-0.5 (0.2), p=0.02). Learning rate was reduced with both anxiety and irritability (b=-0.11 (0.04), p=0.01).

Generalized mixture modeling identified two learning types. Individuals clearly differ in the degree to which they are affected by feedback on the prior trial, even if the stimulus was far away on the morph continuum. This represents generalization. There is a group showing very high generalization and a group with lower generalization.

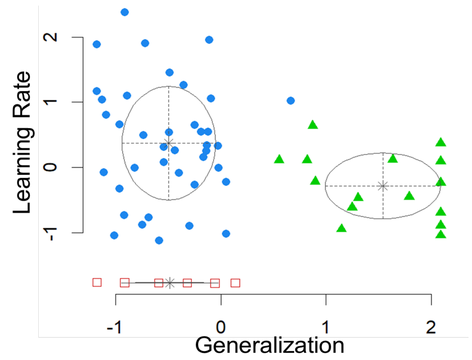


***Figure 3.*** Gaussian mixture model clustering shows 3 types of IBT learners. Learning rate reflects learning ability and generalization reflects how broadly feedback is applied to neighbors on the face-emotion continuum. Units are z-scores. The rightward group represents individuals who learn moderately well and apply feedback broadly. The upper left group has variable learning and applies feedback to few neighboring face stimuli.

## Discussion

The model-based analysis showed that age and affective symptoms modulate category learning during IBT.

Very high generalization suggests a distinctive impairment or learning strategy not accounted for in the prior literature. Lower learning rates with higher negative affectivity suggest an impairment. In an independent RCT of IBT for adult depression, learning rate as measured by our model was positively associated with clinical improvement.

This work represents the first clinical translation of ALCOVE, a prominent learning model. Understanding learning will improve the ability to identify learning types, improving precise prescription of IBT.

The neurocognitive bases of these findings are largely unknown but may be explored in future research towards understanding pathophysiology and treatment targets.

## Acknowledgements

This work was supported by grant number K23MH113731 from the National Institutes of Health/National Institutes of Mental Health.

Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the NIH.

## References

Kruschke, John K. "ALCOVE: An exemplar-based connectionist model of category learning." *Psychological Review* 99.1 (1992): 22-44.
Penton-Voak, Ian S., Marcus R. Munafò and Chung Yen Looi. "Biased facial-emotion perception in mental health disorders: a possible target for psychological intervention?." *Current Directions in Psychological Science.* 26.3 (2017): 294-301.