# Affective and Interpersonal Communication

Advancing an Interdisciplinary Research Program

Jeffrey M. Girard

KU *Psychology*

jmgirard@ku.edu

affcom.ku.edu

# Since Joining KU in 2020…

*Teaching*

4 new course designs
14 full courses taught
293 students enrolled

34 undergraduate RAs
6 honors thesis mentees

5+2 graduate mentees
1 postdoctoral mentee

*Research*

32 new publications
7 new as first author
3,005 new citations

30 grant proposals
10 grants awarded

2 new open databases
14 new open software

*Service*

BBQ Program Director ('22–Now)
KDSC Co-director ('23–Now)
Dean's Faculty Fellow ('24–'25)

SITAR President ('25)
ACII Program Chair ('23)
AFGR Program Chair ('25)

IEEE TAC Associate Editor ('20–'25)
Collabra Associate Editor ('21–Now)

# New and Deepening Collaborations

- Arian Ashourvan
- Monica Biernat
- Tera Fazzino
- Kelsie Forbush
- Nancy Hamilton
- Katie Hoemann
- Steve Ilardi
- Amber Watts

McLean HARVARD MEDICAL SCHOOL AFFILIATE

Google

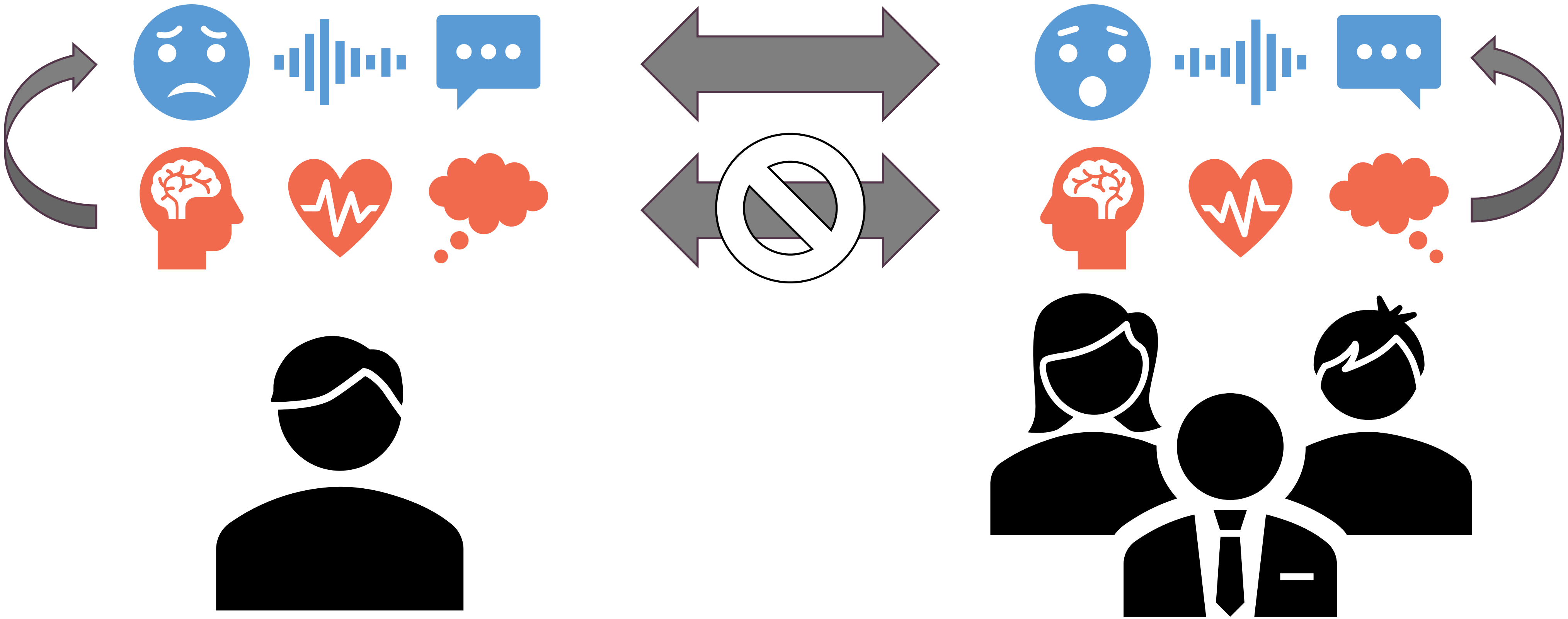
UC Health

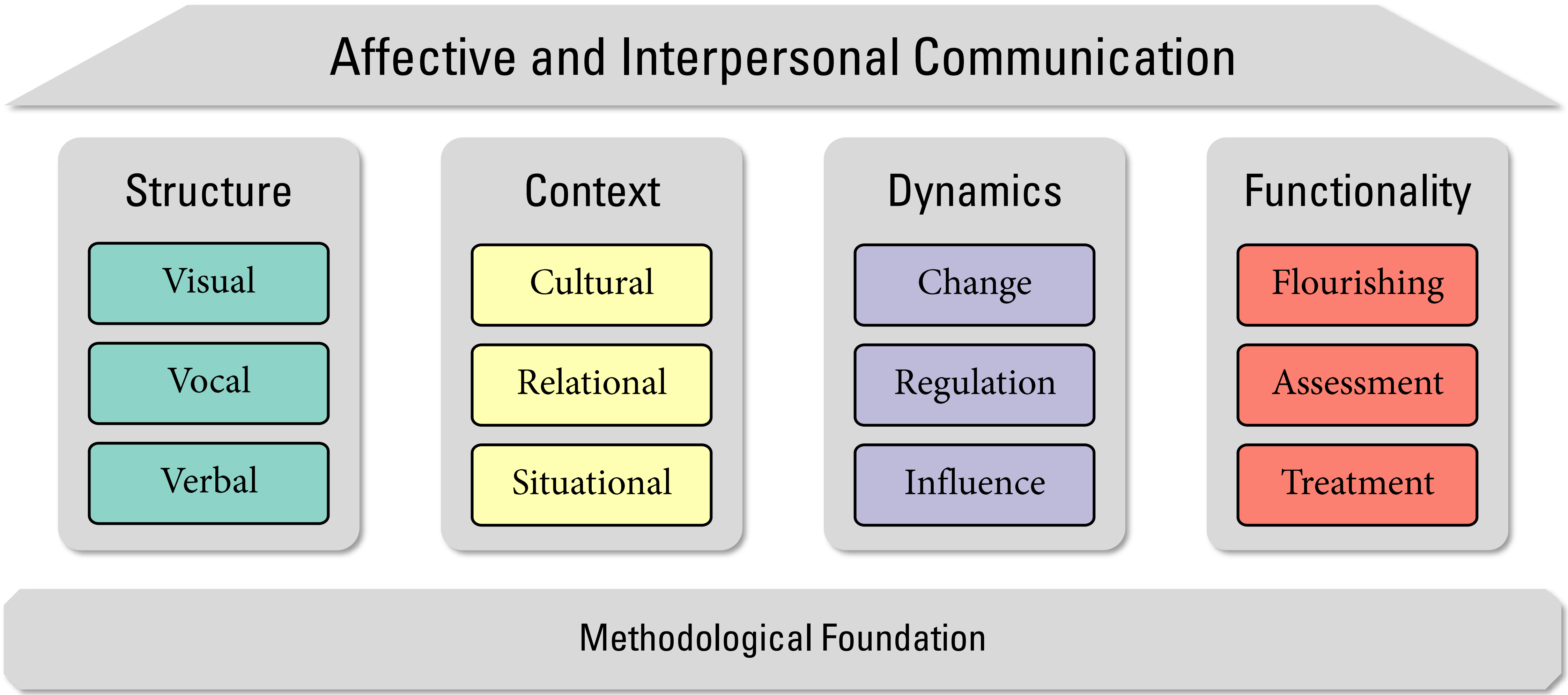
Meta

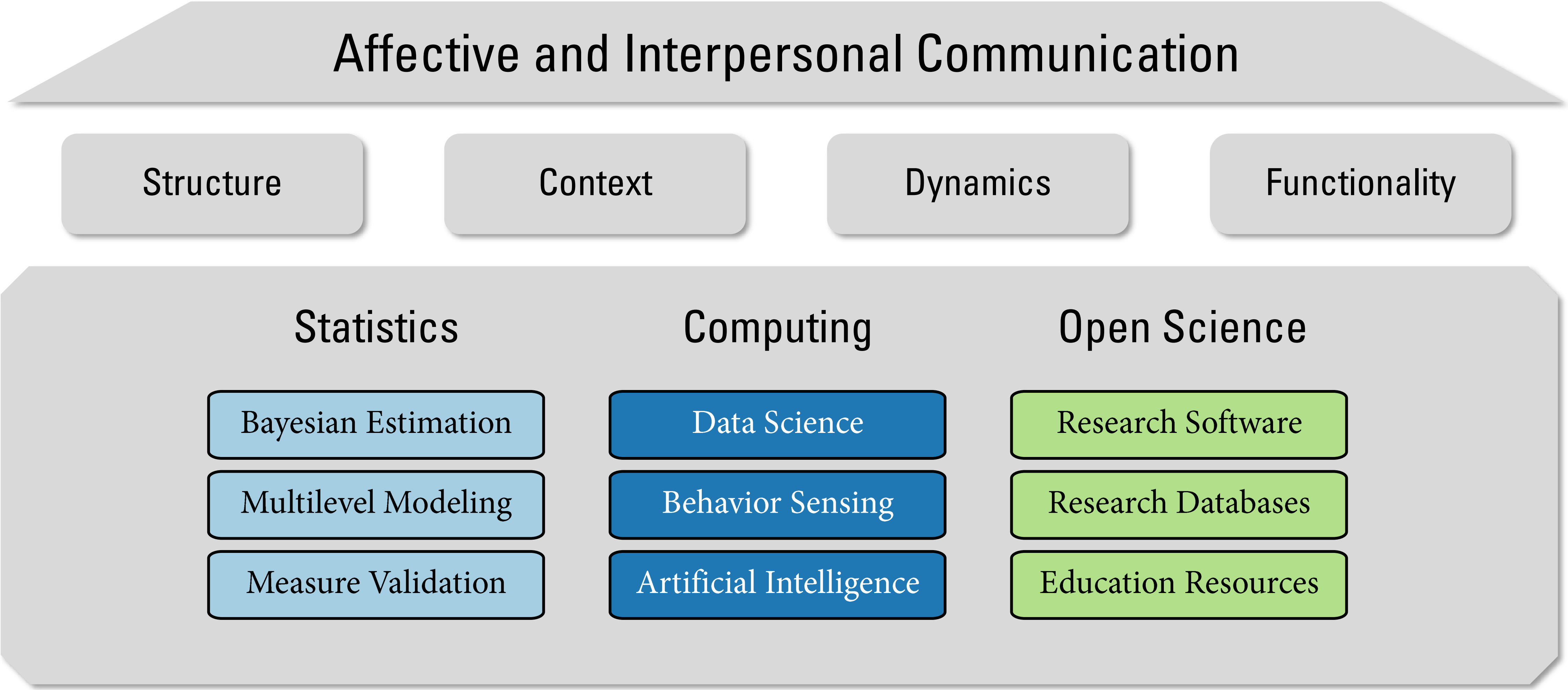
CMU

SMaRT Workshops
Statistics, Methods, and Research Training

# Communicative Behavior Connects Us




Affective and Interpersonal Communication
Structure
Visual
Vocal
Verbal
Context
Cultural
Relational
Situational
Dynamics
Change
Regulation
Influence
Functionality
Flourishing
Assessment
Treatment
Methodological Foundation


Affective and Interpersonal Communication
Structure
Context
Dynamics
Functionality
Statistics
Computing
Open Science
Bayesian Estimation
Multilevel Modeling
Measure Validation
Data Science
Behavior Sensing
Artificial Intelligence
Research Software
Research Databases
Education Resources

# Evolution of the Framework

- **Structure:** Increasing emphasis on verbal communication (i.e., natural language processing)
- **Context:** Developing theoretical and methodological frameworks for studying context
- **Dynamics:** Increasing emphasis on interpersonal dynamics (e.g., dyadic synchrony)
- **Functionality:** Increasing emphasis on transdiagnostic assessment of broader populations

- **Statistics:** Increasing emphasis on generalizability theory and longitudinal modeling
- **Computing:** Increasing emphasis on artificial intelligence (especially validation efforts)
- **Open Science:** Teaching literate programming and containerization for open science

# Highlighted Projects

1. **Dynamic & Dyadic Smiling in Depression Therapy:**
   - This psychological project was enhanced by our computational background
   - *Structure* (Visual), *Dynamics* (Influence), *Functionality* (Treatment)
   - *Statistics* (Multilevel), *Computing* (Behavior Sensing), *Open Science* (Software)

2. **Validating AI for Sentiment Analysis of Natural Speech:**
   - This computational project was enhanced by our psychological background
   - *Structure* (Verbal), *Context* (Cultural), *Functionality* (Assessment)
   - *Statistics* (Validation), *Computing* (Artificial Intelligence), *Open Science* (Software)

# Dynamic & Dyadic Smiling in Depression Therapy

Girard, Yermol, Bylsma, Cohn, Fournier, Morency, & Swartz

*Journal of Consulting and Clinical Psychology* (2025)

# Background and Rationale

- The working alliance is the collaborative and emotional bond between patient and therapist
- Dyads with stronger working alliances tend to have better outcomes, across types of therapy
- **What behavioral factors contribute to the strength of the working alliance?**
- Koole & Tschacher's (2016) "In-Sync" Model of Psychotherapy theorizes:


- Interpersonal Synchrony is the coordination of individuals' behaviors or states over time
- **Why have previous findings linking synchrony to alliance (and outcomes) been *mixed?***

# Possible Explanations

1. **What behavior is being synchronized?**
   - We examine specific actions: smiles (and scowls)
2. **Is it really synchrony or just occurrence?**
   - We partialize behavior synchrony and occurrence
3. **Who is rating the working alliance?**
   - We compare patient-ratings and therapist-ratings
4. **Are these state-like or trait-like processes?**
   - We decompose within- and between-dyad effects

# Data and Measures

- 65 outpatients with a DSM-5 depressive disorder (65% female, 68% white, 32 ± 14 years)
- 14 therapists with 1 to 25 years experience (80% female, 87% white, 35 ± 9 years)
- Patients were randomized to receive 8 sessions of either brief CBT or brief IPT

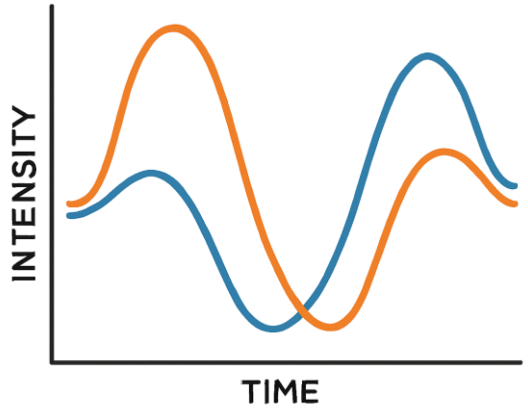
| *Before* Each Session | *During* Each Session | *After* Each Session |
| --- | --- | --- |
| **QIDS-16**<br>(Patient Self-report Depression Severity) | **Smile (and Scowl) Intensity**<br>(Measured for both Patient and Therapist, Computer Vision and Machine Learning, 0.1s Bins, Validation $r$ = 0.85 (and 0.70) | **WAI-SR**<br>(Patient and Therapist Report Strength of Working Alliance) |

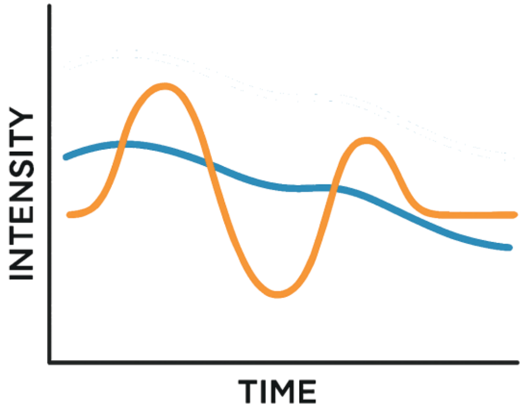
- 65 dyads × 2 people/dyad × 8 sessions × 1 hour/session × 36,000 bins/hour ≈ 37.44M bins

# Quantifying Synchrony

- Interpersonal synchrony was measured using windowed cross-correlation (WCC)
  - How linked are two time series within short windows and across different lags?
  - Yields one number representing the strength of patient–therapist synchrony per session

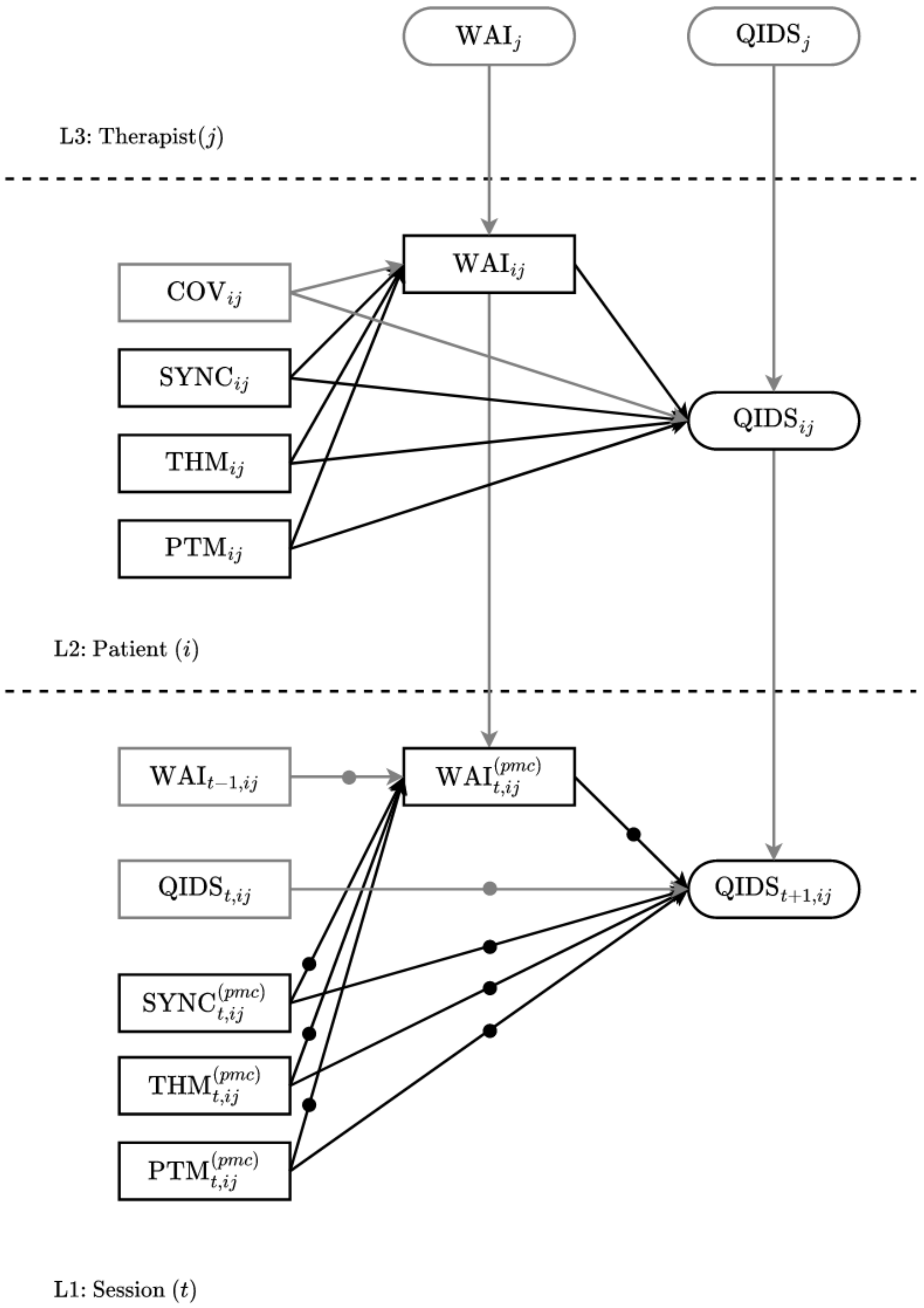
*Higher* WCC Example



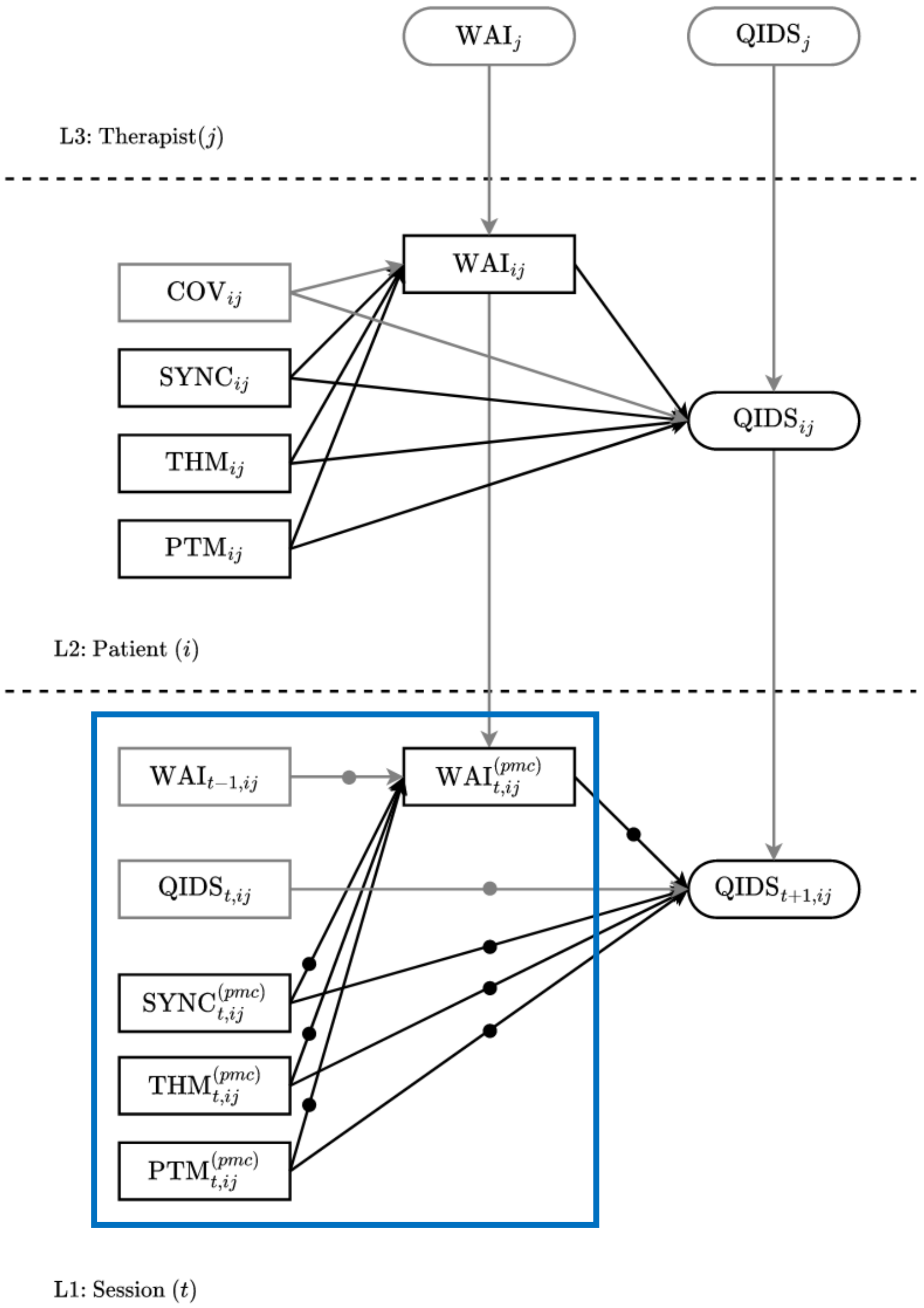

*Lower* WCC Example

  - Developed a new R package (github.com/jmgirard/wcc) for efficient, parallelized computation

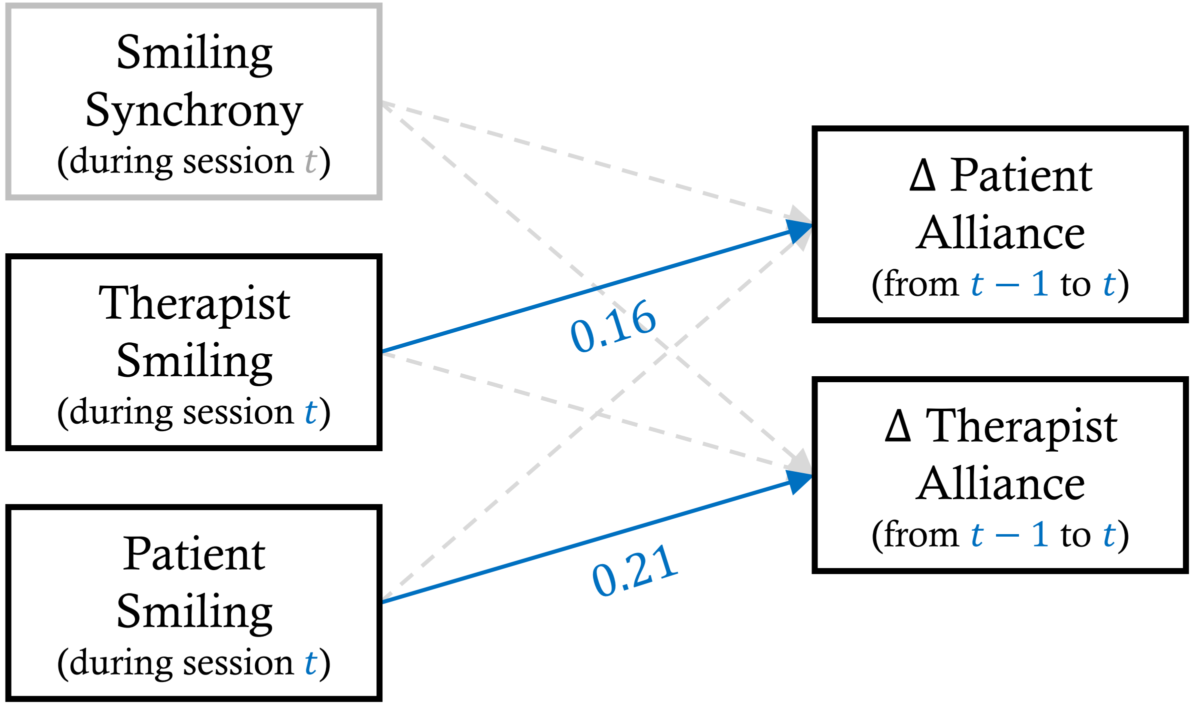



# ANALYTICAL STRATEGY

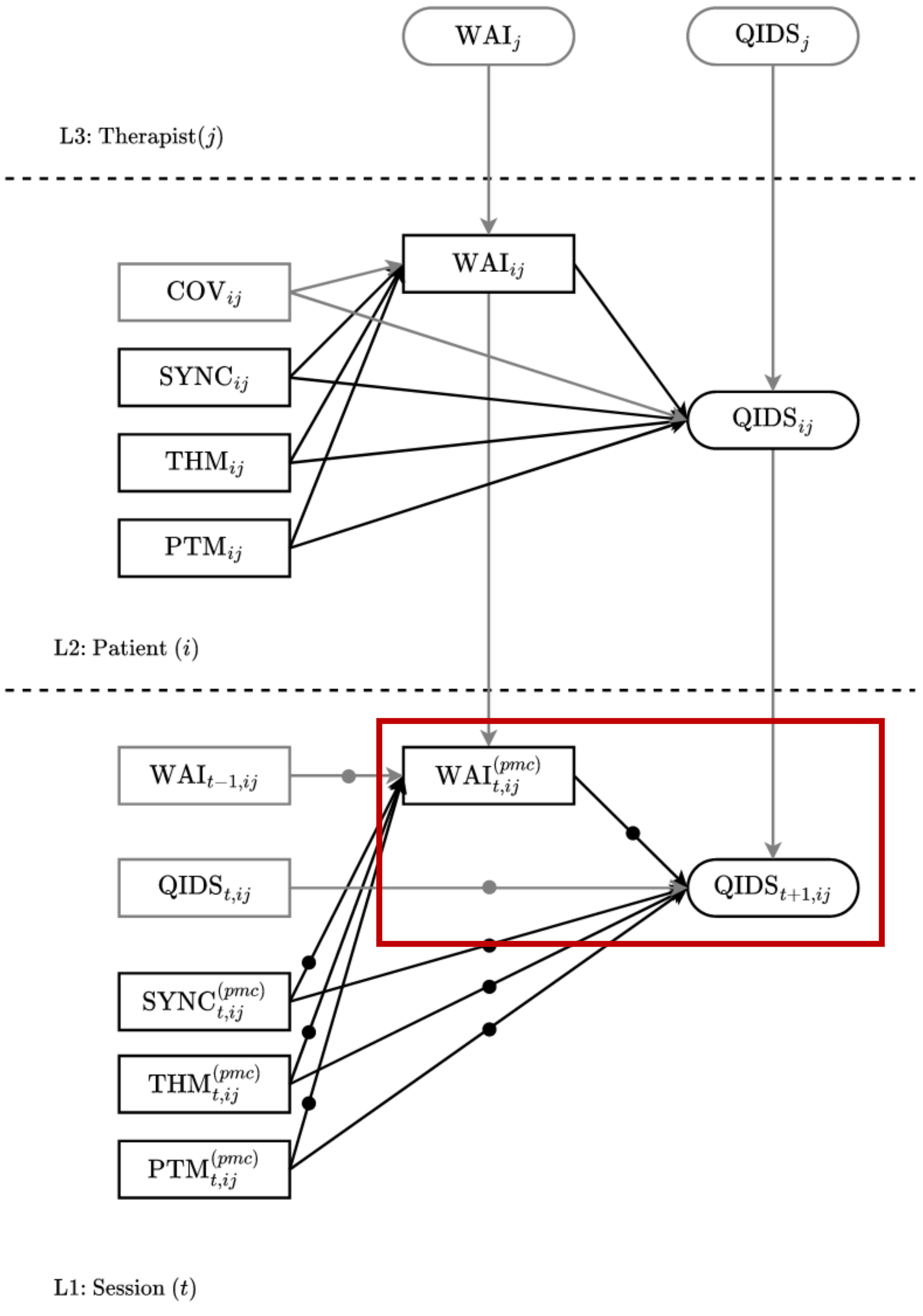
- Hierarchical Bayesian Mediation Modeling
- Sessions $(t)$ in Patients $(i)$ in Therapists $(j)$
- Behavior → Alliance → Depression Mediation
- Both Within-dyad and Between-dyad Levels
- Controls for Patient and Therapist Occurrence
- Separate models for Smiling vs. Scowling
- Separate models for Patient vs. Therapist WAI



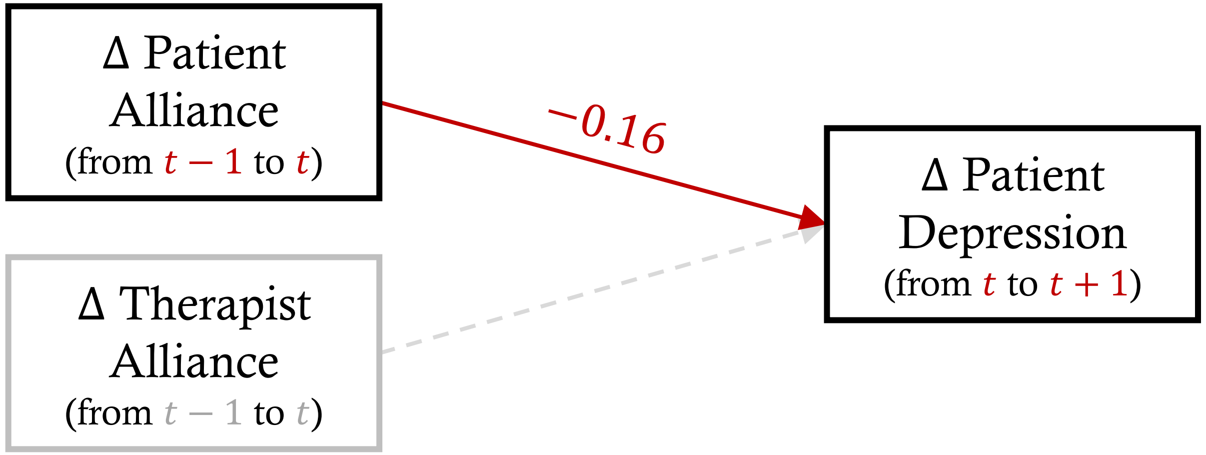

# Smiling, Within-Dyad



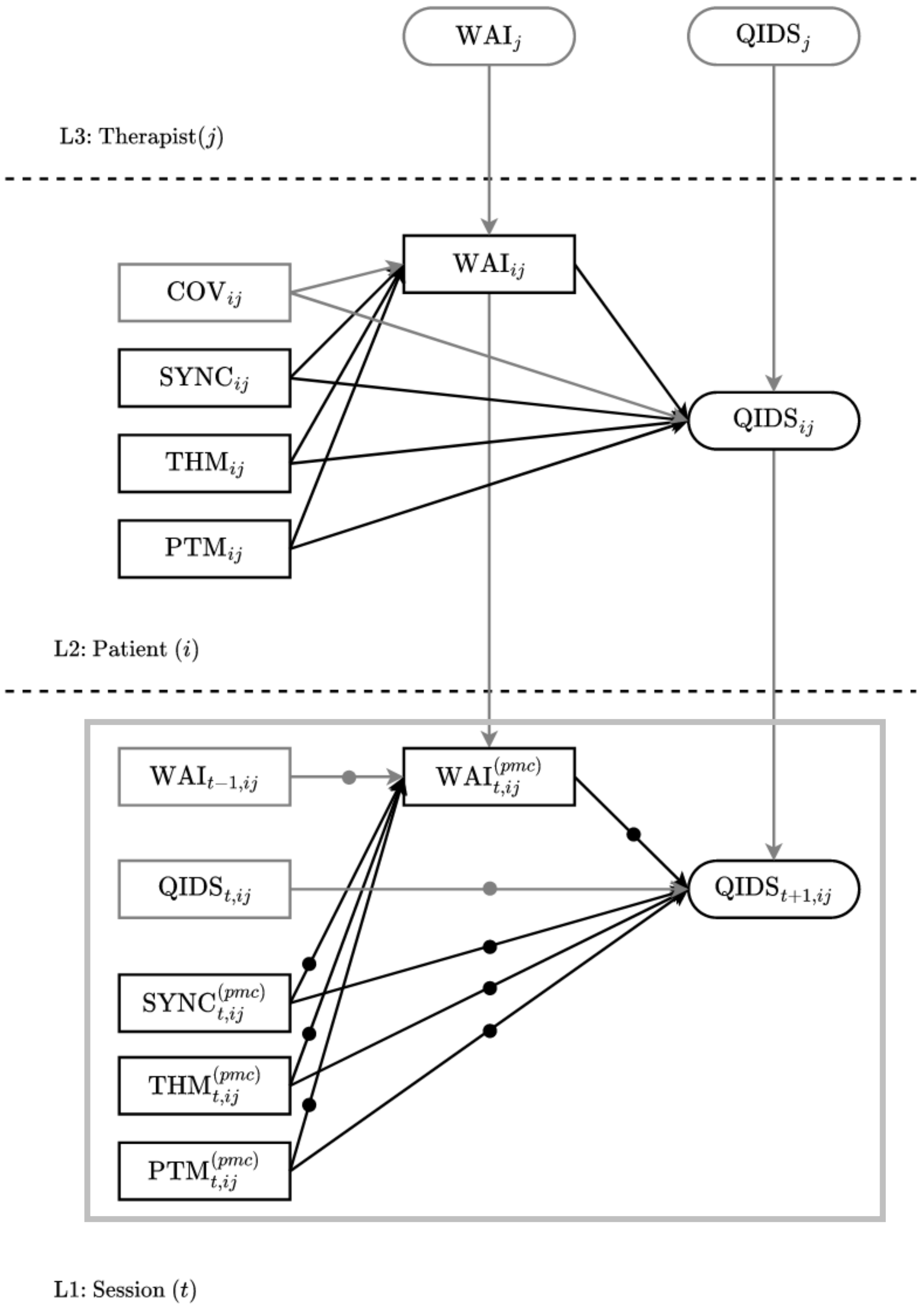

*Note.* Grey, dashed pathways are not significant.



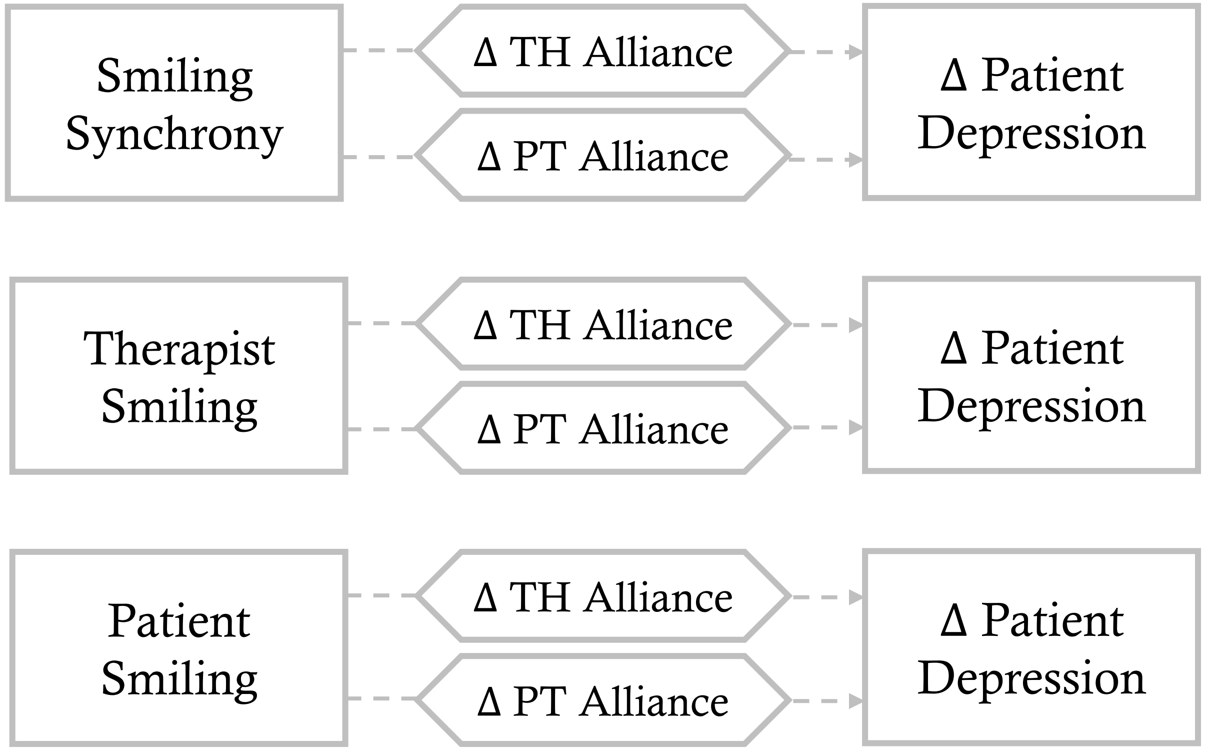

# ALLIANCE, WITHIN-DYAD



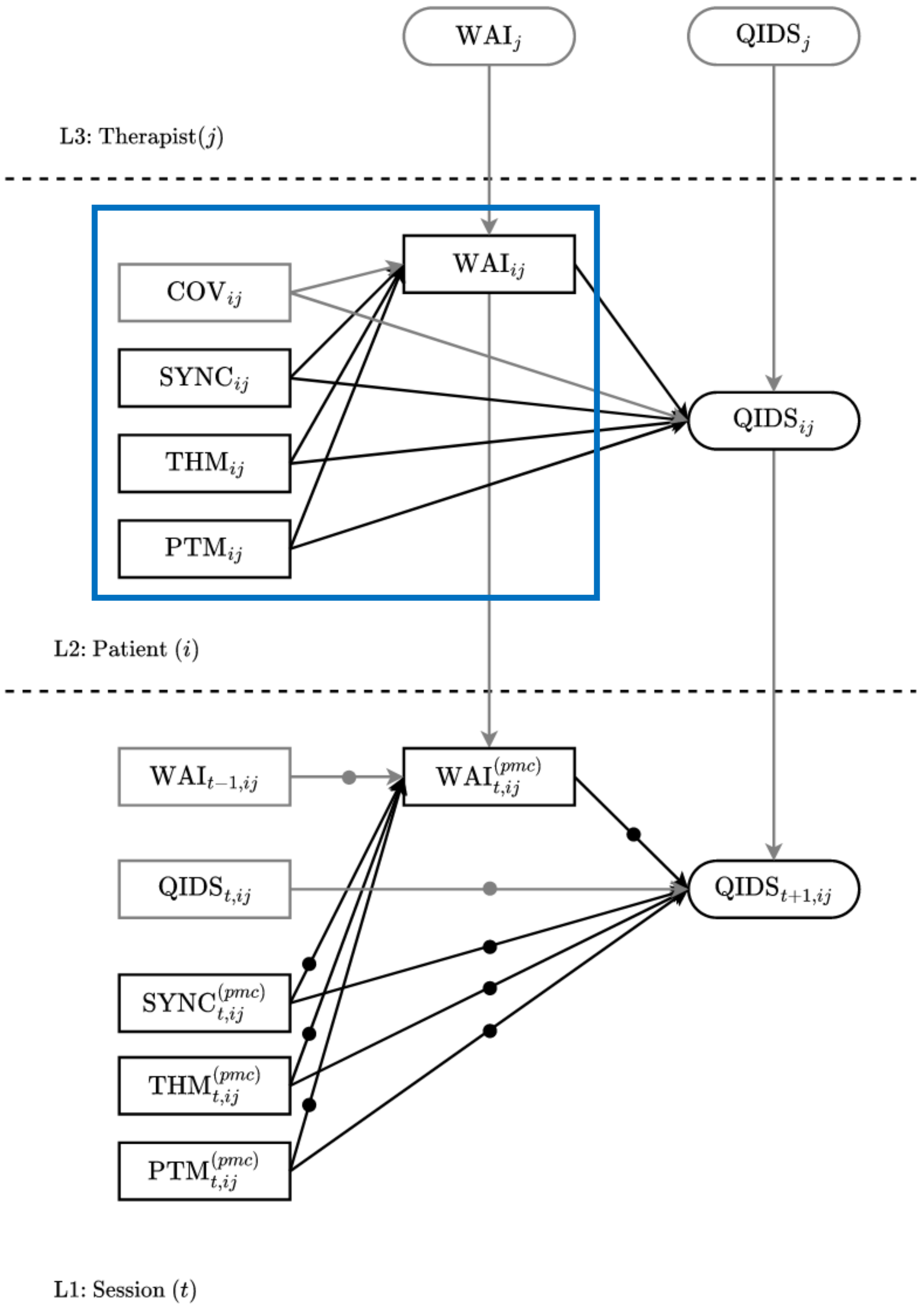

*Note.* Grey, dashed pathways are not significant.



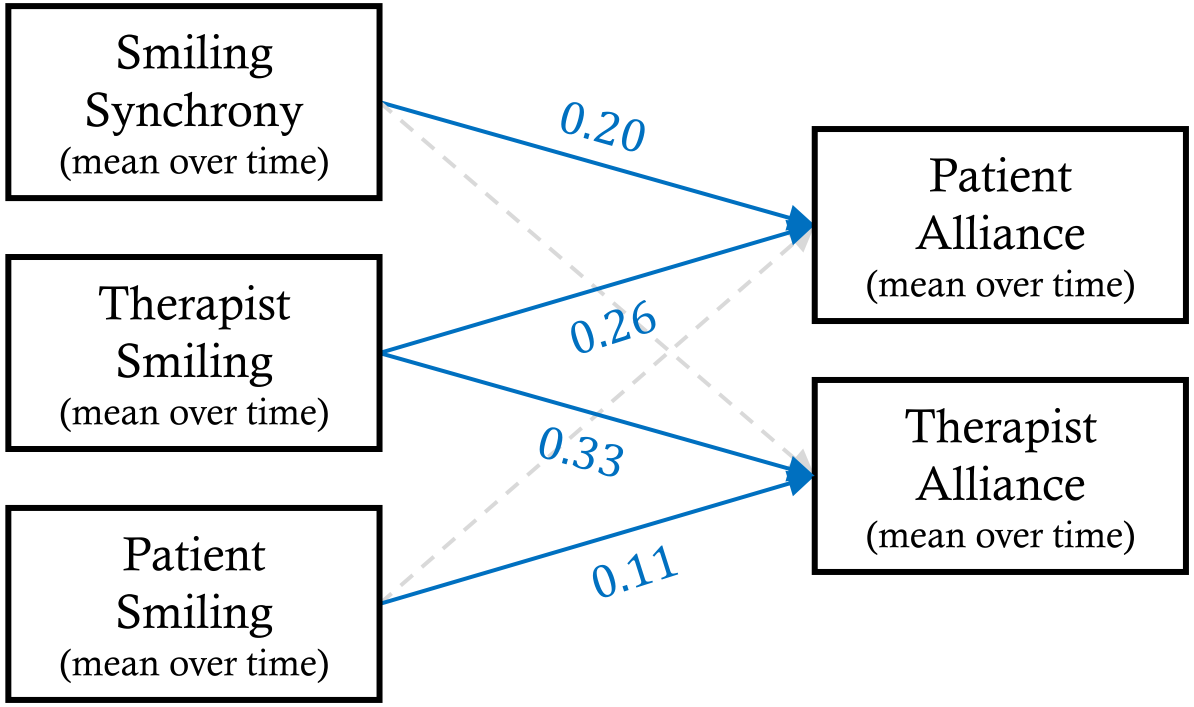

# WITHIN-DYAD MEDIATION



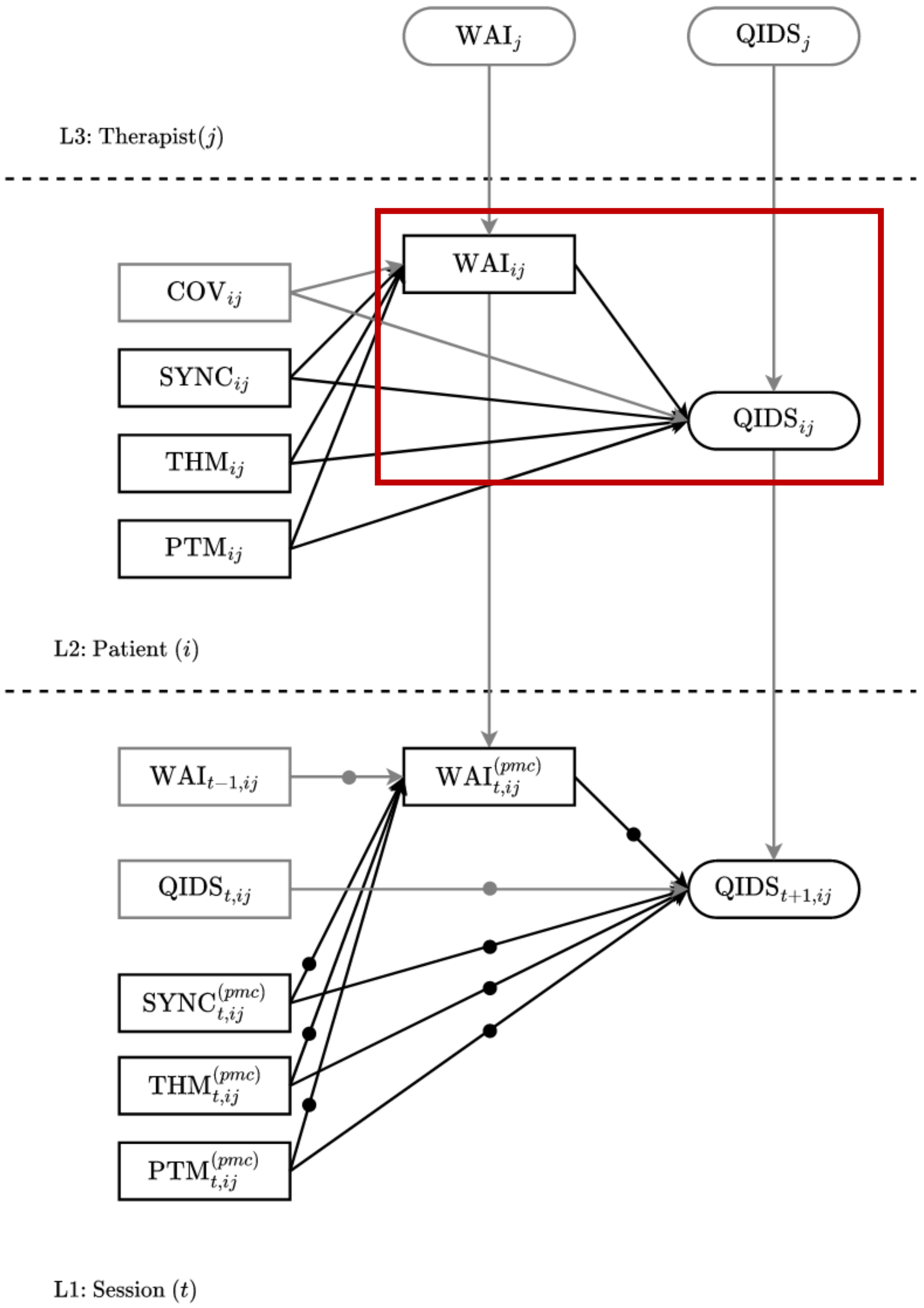

*Note.* Grey, dashed pathways are not significant.



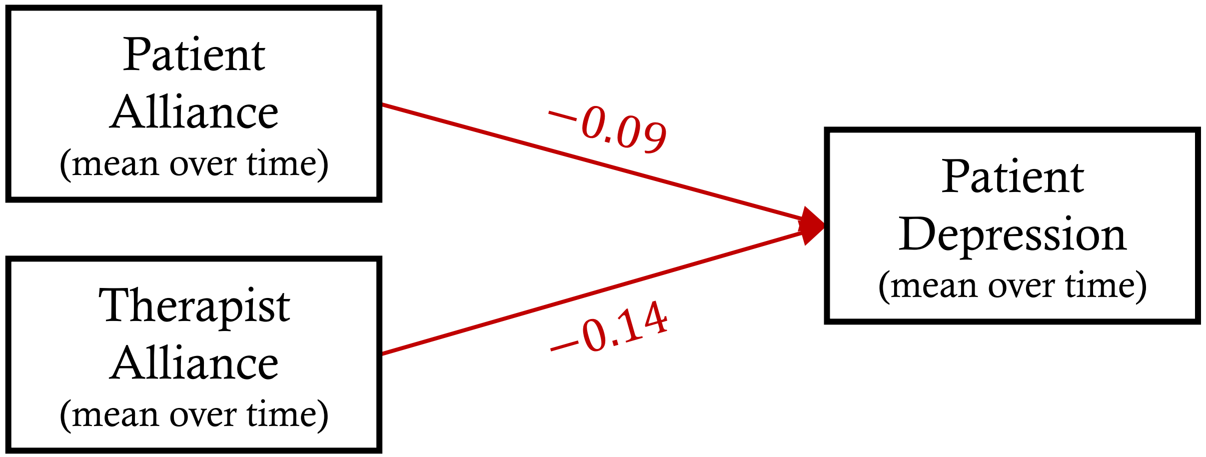

# SMILING, BETWEEN-DYAD

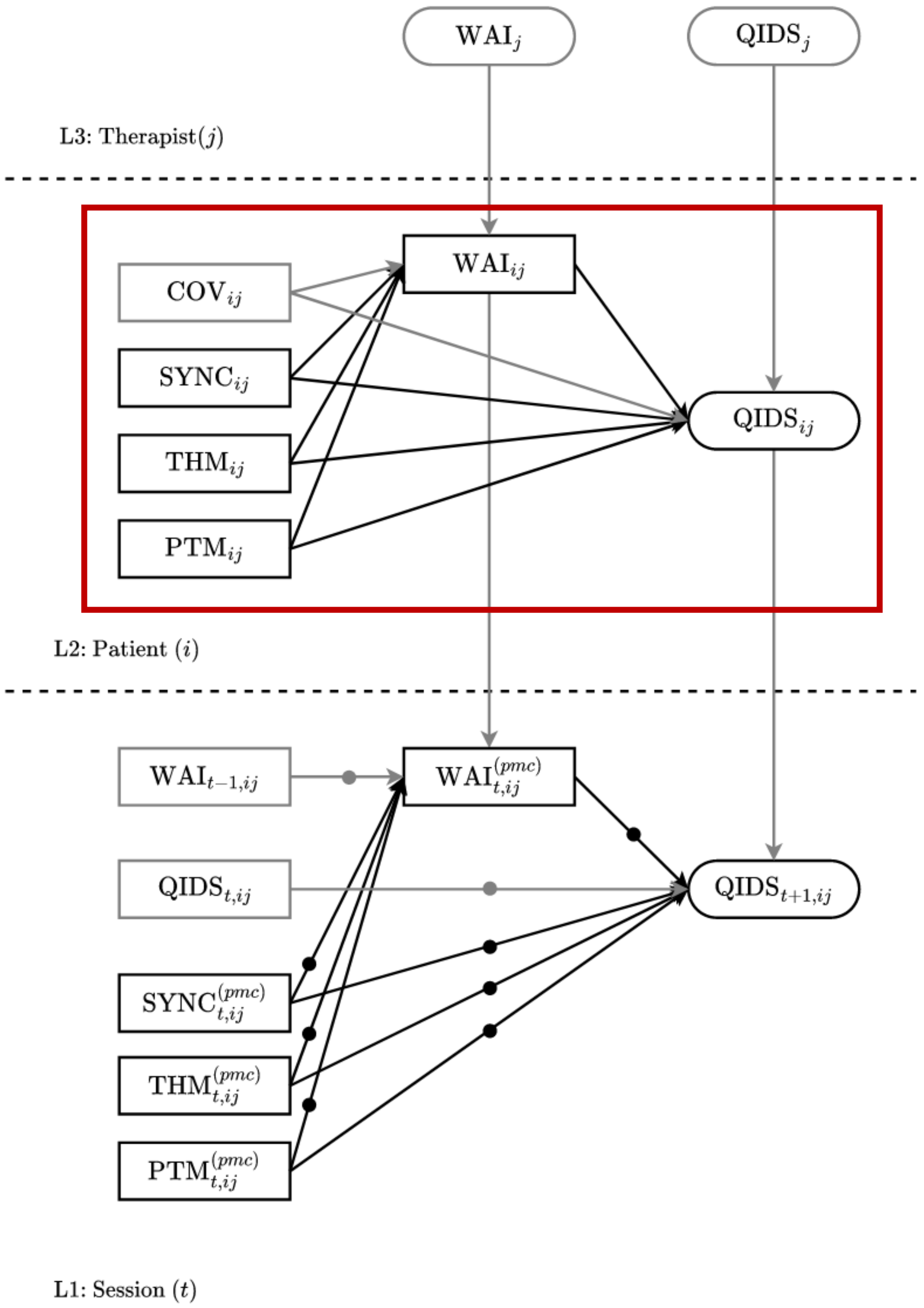



*Note.* Grey, dashed pathways are not significant.

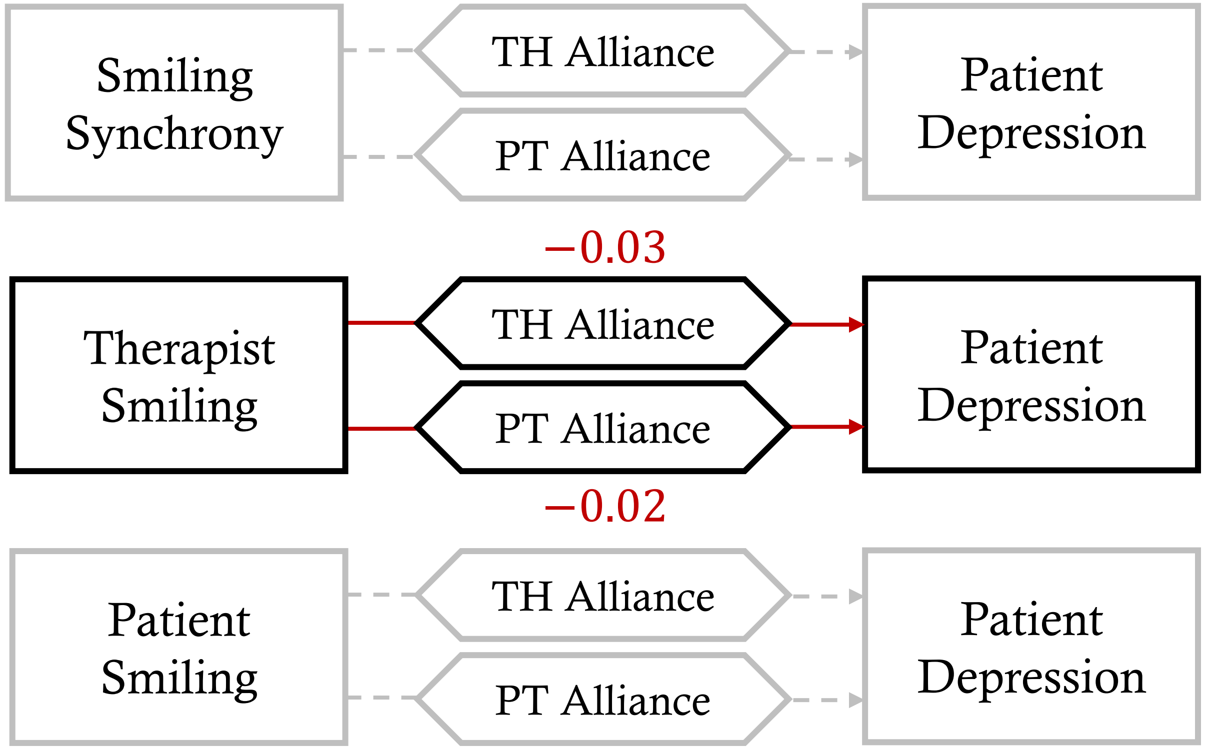



# ALLIANCE, BETWEEN-DYAD






# Between-Dyad Mediation




*Note.* Grey, dashed pathways are not significant.

# Conclusions

- Patients and therapists use their partner's smiling in a session to update their alliance ratings
- Changes in patient (but not therapist) alliance ratings predict subsequent changes in symptoms

- Dyads with stronger patient alliance ratings had more therapist smiling and smile synchrony
- Dyads with stronger therapist alliance ratings had more therapist smiling and patient smiling
- Alliance mediated the *between-dyad* relationship between therapist smiling and outcomes

- Measuring momentary synchrony was only possible due to *behavior sensing* techniques
- These nuanced insights were only possible to derive due to sophisticated *statistical* techniques

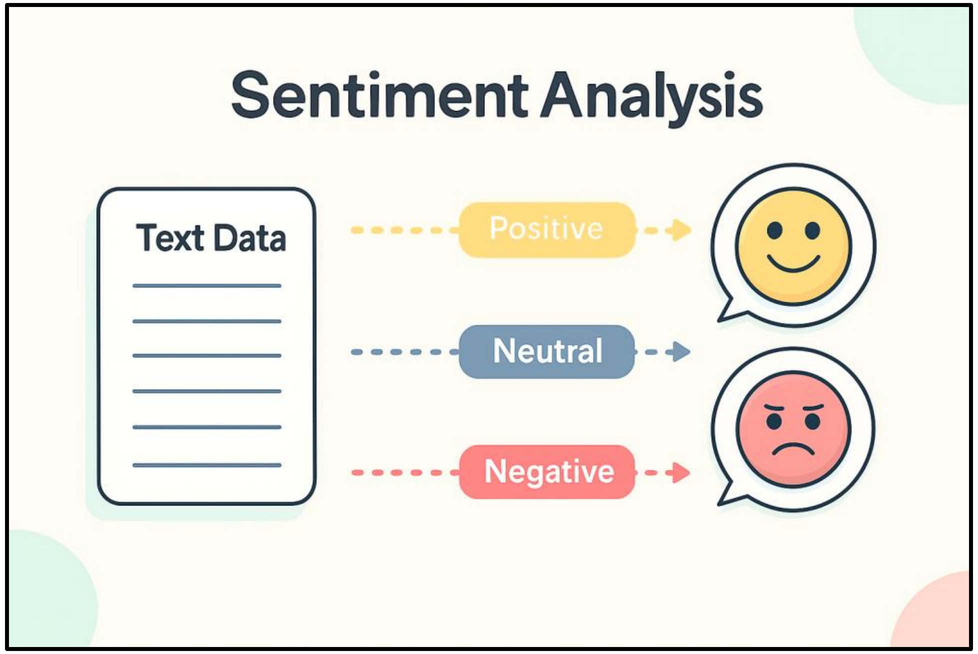
# Validating AI for Sentiment Analysis of Natural Speech

Girard, Jun, Ong, Liebenthal, & Baker

*Affective Science* (R&R)

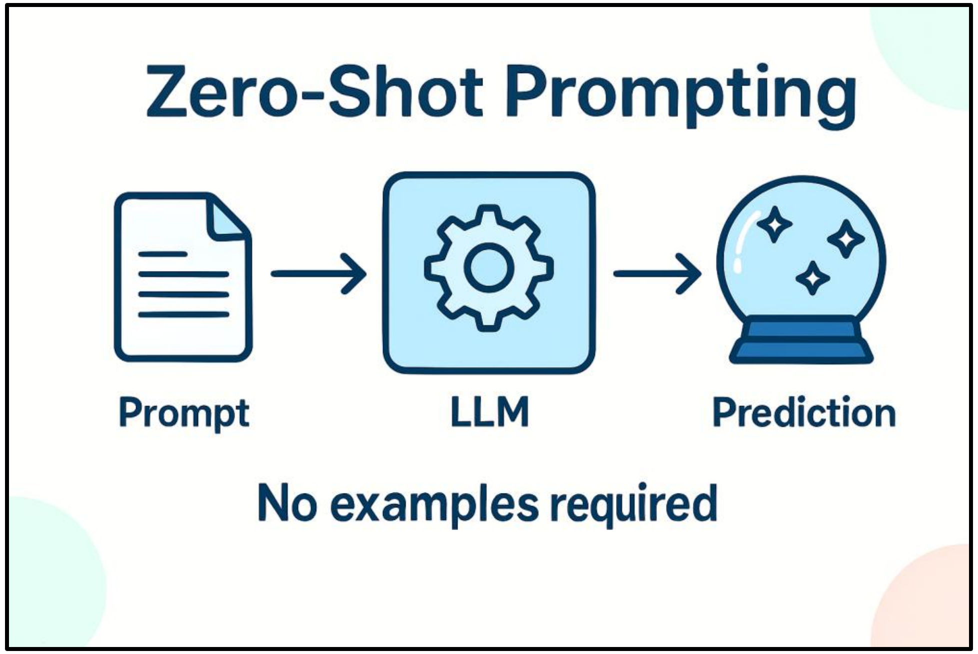
# Background and Rationale

- Sentiment analysis identifies the **emotional** or **evaluative tone** of language (e.g., positive to negative)
- Rule-based approaches (e.g., LIWC) count positive and negative words using human-created dictionaries
- These simple methods often fail to capture context, sarcasm, allusions



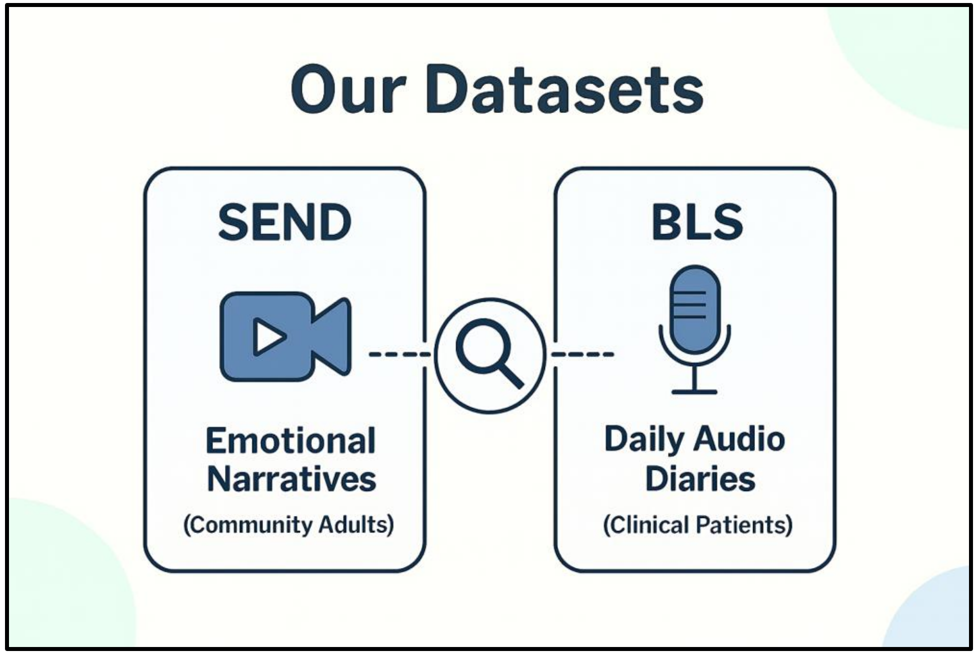
# Research Questions

- Large Language Models (LLMs) are AI systems that can better capture context and meaning in text data
- Open-weight LLMs can run locally, privately, securely (not in the cloud)
- **Can open-weight LLMs estimate sentiment without extra training?**
- **Are such models *fair* across groups?**



# Data and Measures

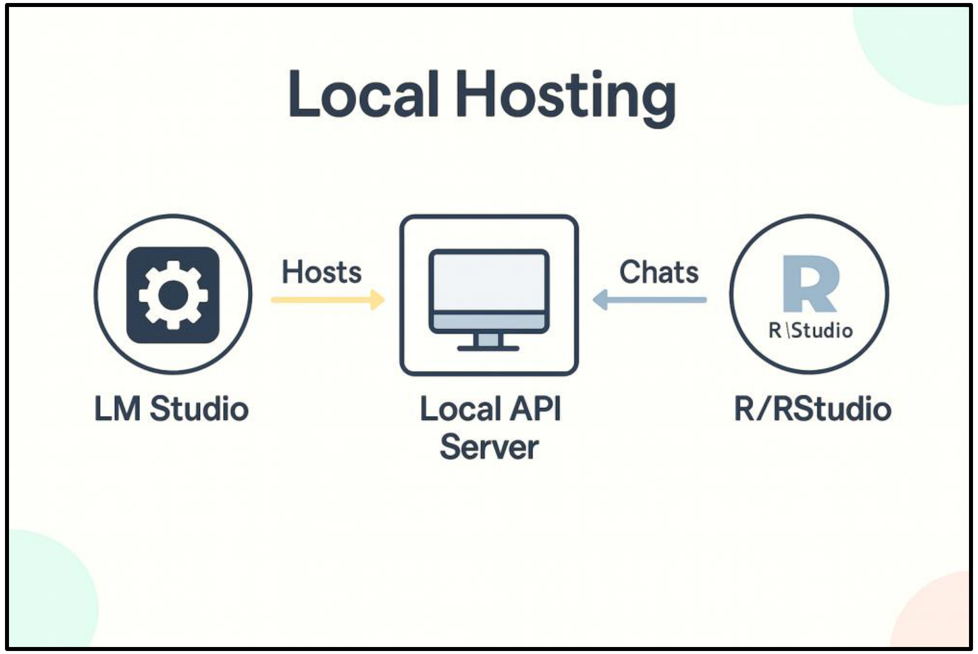
- **SEND** (Stanford Emotional Narratives Dataset)
  193 life narrative videos from
  49 students and community adults
  Ratings from 19 human observers
- **BLS** (Bipolar Longitudinal Study)
  297 daily audio journals from
  64 patients with bipolar disorders
  Ratings from 5 human observers
- Compare LLM predictions to ratings



# Infrastructure and Software

- Deployed 20 open-weight LLMs (1B-120B parameters) via LMStudio
- Used a consumer GPU workstation with an NVIDIA RTX 4060 (~$350)
- LMStudio hosts a local server with API access via our new R package

github.com/affcomlab/easyllm



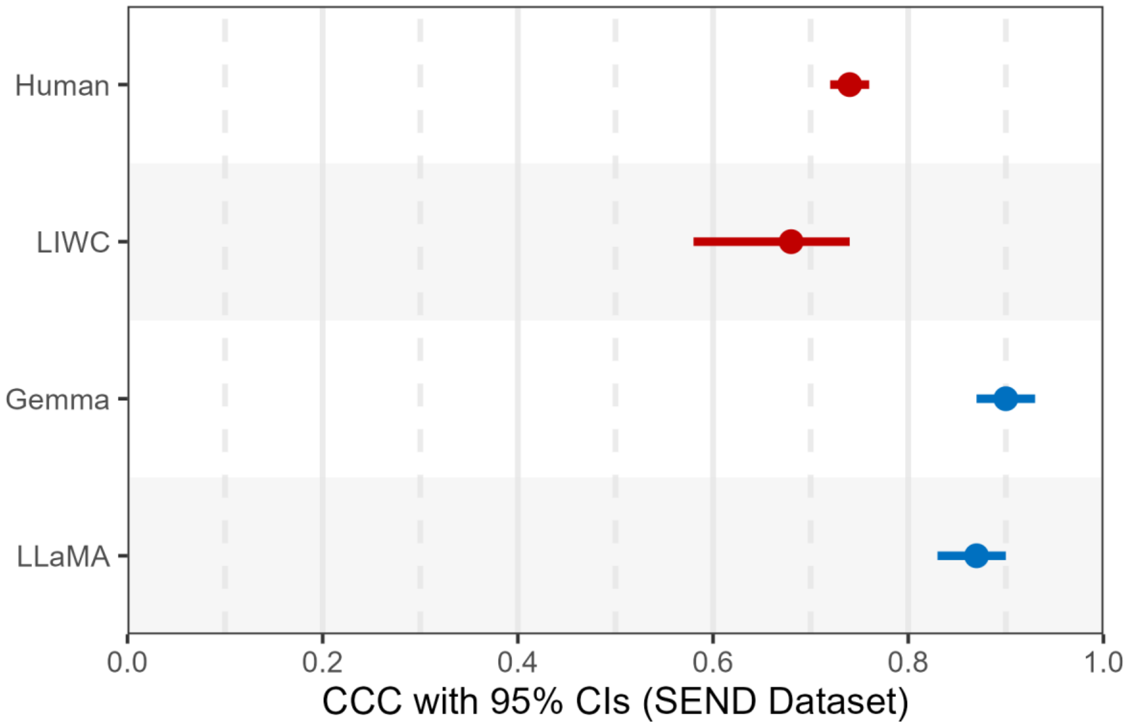
# Prompt Engineering

- **User prompt:** contains a transcript to perform sentiment analysis on
- **System prompt:** defines the task, the rating scale, and jailbreaks
- In R, we loop through transcripts, create prompts, and capture output
- Each transcript is rated separately

```
You will be provided with
a transcript, and your
task is to rate its
sentiment on a scale from
1 to 7 where 1 represents
'very negative' and
7 represents
'very positive'...
```

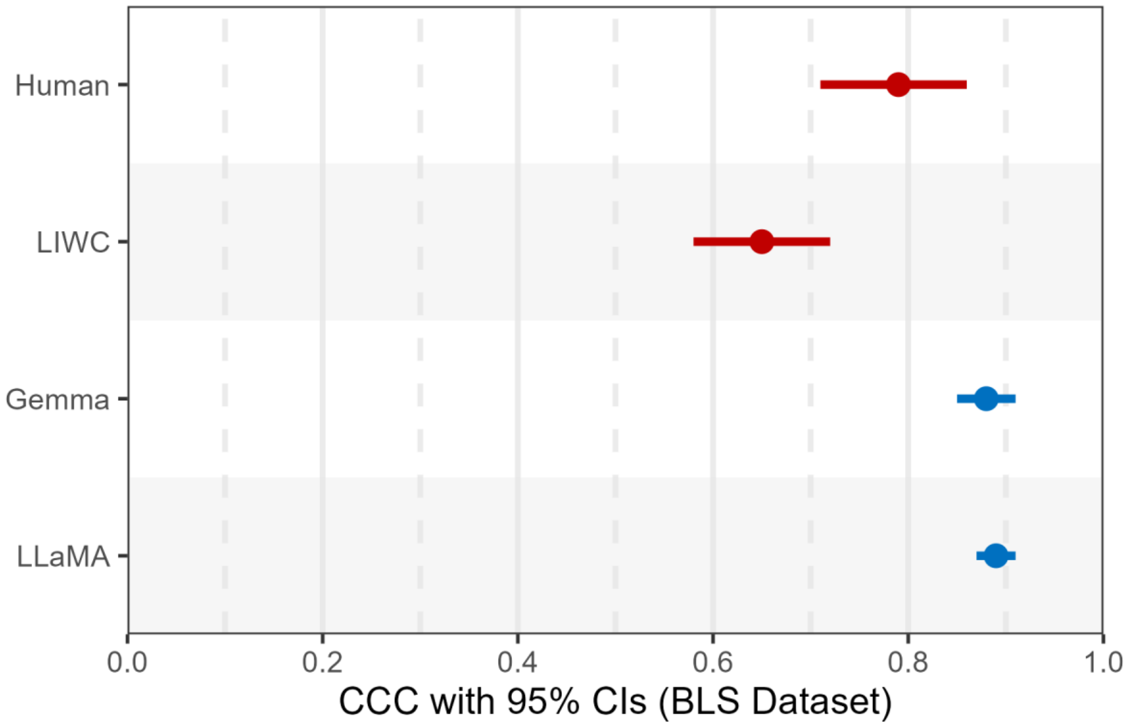
# Results on Non-clinical SEND Dataset

- **Benchmarks**
  - Human baseline (.74)
  - LIWC-22 baseline (.68)
- **Highlighted LLMs**
  - Google Gemma 3–12b (.90)
  - Meta LLaMA 3.3–70b (.87)



# Results on Clinical BLS Dataset

- **Benchmarks**
  - Human baseline (.79)
  - LIWC-22 baseline (.65)
- **Highlighted LLMs**
  - Google Gemma 3–12b (.88)
  - Meta LLaMA 3.3–70b (.89)



# New Approach to Fairness Analysis

- Adapt a Bayesian cumulative link mixed-effects location-scale model
- Compare sex, race, education, language, and diagnostic groups
- **Bias** = predictions too high or low
- **Sensitivity** = predictions are less strongly related to real changes
- **Consistency** = predictions are less precise (noisier or more variable)

$$y_{ij} \sim \text{Cumulative}(\tau, \eta_{ij}, \delta_{ij})$$

$$\eta_{ij} = \beta_{0j} + \beta_{1j} x_{ij} + \sum_{m=1}^{M} \left( \boldsymbol{\beta}_{\mathbf{2m}} g_{mj} + \boldsymbol{\beta}_{\mathbf{3m}} x_{ij} g_{mj} \right)$$

$$\beta_{0j} = \gamma_{00} + u_{0j} \qquad \beta_{1j} = \gamma_{10} + u_{1j}$$

$$\begin{bmatrix} u_{0j} \\ u_{1j} \end{bmatrix} \sim \mathcal{N}\left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_0^2 & \rho\sigma_0\sigma_1 \\ \rho\sigma_0\sigma_1 & \sigma_1^2 \end{bmatrix} \right)$$

$$\log(\delta_{ij}) = \alpha_0 + \sum_{m=1}^{M} \boldsymbol{\alpha}_{\mathbf{1m}} g_{mj}$$

Table 2: Focal Results from the Fairness Analysis Models

| | Bias ($\beta_{2m}$) | | Sensitivity ($\beta_{3m}$) | | | Consistency ($\alpha_{1m}$) | | |
|---|---|---|---|---|---|---|---|---|
| | Estimate | $p$-value | Estimate | $p$-value | | Estimate | $p$-value | |
| SEND Dataset: Gemma 3 (12b) | | | | | | | | |
| 1. Sex (Male – Female) | 0.12 | .318 | −0.57 | .088 | † | 0.05 | .418 | |
| 2. Race (Asian – White) | 0.23 | .399 | 1.57 | .054 | † | −0.04 | .440 | |
| Race (Black – White) | 0.59 | .394 | 2.21 | .206 | | −0.46 | .173 | |
| Race (Hispanic – White) | 1.19 | .223 | 0.33 | .412 | | −0.14 | .377 | |
| Race (Multiracial – White) | −0.49 | .279 | 0.81 | .176 | | 0.51 | .057 | † |
| 3. Education level (ordinal) | −0.10 | .387 | −0.36 | .120 | | 0.26 | .044 | * |
| 4. Language (Other – English) | 0.45 | .240 | −0.11 | .430 | | −0.19 | .220 | |
| BLS Dataset: LLaMA 3.3 (70b) | | | | | | | | |
| 5. Sex (Male – Female) | −0.10 | .337 | −0.39 | .067 | † | 0.24 | .057 | † |
| 6. Race (Asian – White) | 0.37 | .288 | −0.26 | .339 | | −0.53 | .004 | * |
| Race (Black – White) | 0.13 | .410 | 0.14 | .427 | | −0.40 | .028 | * |
| Race (Other – White) | −0.07 | .467 | 0.78 | .150 | | −0.27 | .230 | |
| 7. Education level (ordinal) | −0.10 | .268 | 0.03 | .427 | | −0.04 | .327 | |
| 8. Diagnosis (MDD – BSD) | 0.11 | .413 | 0.18 | .352 | | −0.09 | .327 | |
| Diagnosis (SSD – BSD) | 0.22 | .268 | −0.22 | .284 | | −0.07 | .320 | |

*Note*: Each numbered set of rows is a separate model. MDD = Major depressive disorder, BSD = Bipolar spectrum disorder, SSD = Schizophrenia spectrum disorder, $^{\dagger}p < .10$, $^{*}p < .05$.

# Fairness Results

No issues with bias or sensitivity

Gemma's consistency in SEND was worse for participants with fewer years of formal education

LLaMA's consistency in BLS was worse for Asian and Black (compared to White) participants

# Conclusions

- The best LLM models were significantly better than both benchmarks
  - They are better than rule-based approaches (LIWC) and a random, single human rater
- Our new fairness analyses revealed detailed and actionable insights
  - These findings can calibrate our trust in the models and guide corrective efforts
  - This approach was influenced by psychometric and social perception theories
- **LLMs offer private and accurate sentiment analysis on naturalistic speech**
  - We are applying this method (with great success) to study mood, psychosis, and trauma

# Next Five Years

- Develop frameworks for studying context
- Examine behavioral patterns transdiagnostically
- Test and extend the use of zero-shot prompting for psychological and clinical applications
- Help industry partners build psychological knowledge into newest AI models and agents
- Deepen connections with industry, establishing career paths for KU students and research funding
- *Thanks to everyone who supported me along the way!*

affcom.ku.edu

## Graduate Student, Postdoctoral*, and External† Mentees

Kassy Gray (KU)

Aaron Simmons (KU)

Yuanyuan Yang (KU)

Dasha Yermol (KU)

Daiil Jun (KU)

Dr. Brett Welch* (UH)

Victoria Lin† (CMU)

Youssouf Kebe† (CMU)

# THANK YOU! QUESTIONS?