INTENSIVE LONGITUDINAL MODELING OF BIG SOCIAL MEDIA DATA

Jeffrey M. Girard, PhD – *University of Kansas*

psych.ku.edu
affcom.ku.edu

PRESENTATION OUTLINE

Study 1

- Predict the follower growth of Instagram influencers
- Log-linear growth curve model fit in mixed effects framework

Study 2

- Compare behavioral rhythms between three time periods
- Dynamic SEM with moderated temporal (periodic) effects

LOG-LINEAR GROWTH MODELING | 1

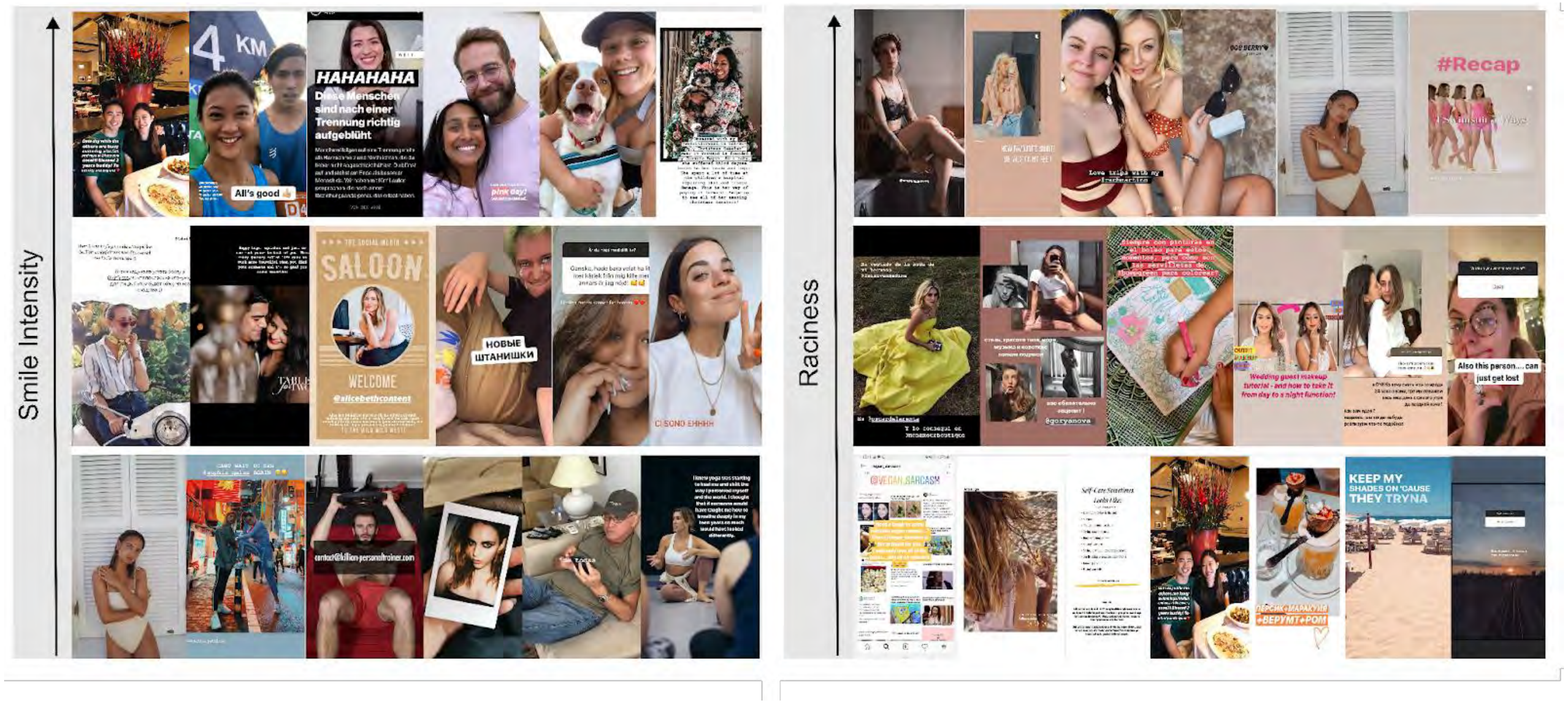
RESEARCH QUESTION

- Instagram influencers are impactful marketing agents around the world
  - Annual spend on influencer marketing was $16.4B in 2022 with 45% from Instagram
- Marketing scholars and agencies are interested in influencer effectiveness
  - What strategies and behaviors promote follower growth and follower spending?
  - How does the growth of an influencer’s following influence consumer engagement?
- **Which influencer behaviors are associated with long-term growth?**
  - *Posting patterns:* how much and which types of content are uploaded?
  - *Captioning:* what is included in the text accompanying uploads?
  - *Personage:* who and what is depicted in the uploaded images?

DATA COLLECTION AND PROCESSING

- Partnered with an Influencer Management Agency (whalar.com)
- Access internal database of Instagram influencers
  - Data provided by 5,835 Whalar-managed influencers across 55 countries
  - All 14.2M content uploads from May 2019 to October 2021were analyzed
- Influencers self-reported their gender, age, and country-of-origin
- Instagram API gave follower count and content type, format, and ad-status
- Text content was analyzed for length, hashtags, mentions, and emojis
- Image content was analyzed for faces, smiles, and raciness (sexual content)

DATA PROCESSING (EXAMPLES)



DATA ANALYSIS CHALLENGES AND SOLUTIONS

Longitudinal Challenges

- Growth is *relative* rather than absolute
- Log-linear growth model on monthly counts

Structural Challenges

- Observations within influencers within countries
- Three-level mixed-effects model

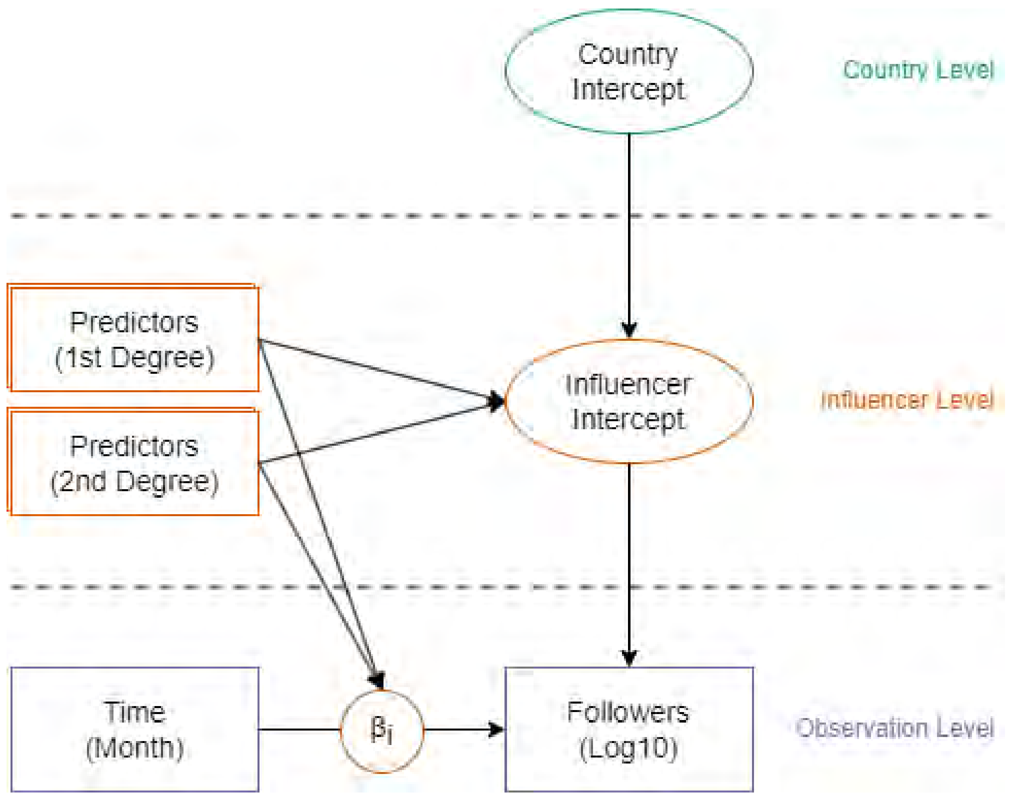
Relational Challenges

- Predictors may have "optimal" values
- Quadratic orthogonal polynomial terms

Visualization Challenges

- We primarily care about predictors of *growth*
  - But growth is not the outcome variable…
  - Rather, it is captured by time moderation
  - Estimate time slopes and map them to the y-axis
- The original units are meaningful/interesting
  - But the variables have been transformed…
  - DV is log-transformed, and IVs are standardized
  - Extract estimates and back-transform them all

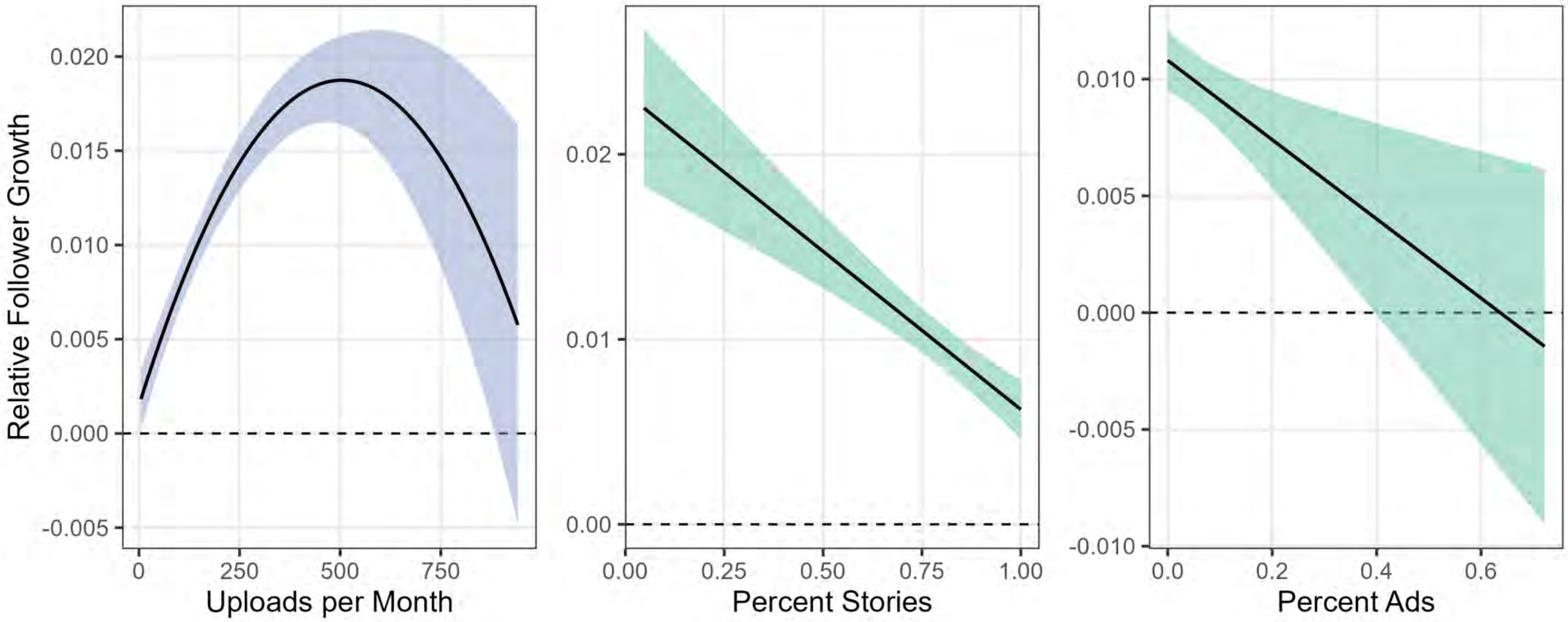
THREE-LEVEL NESTED MODEL



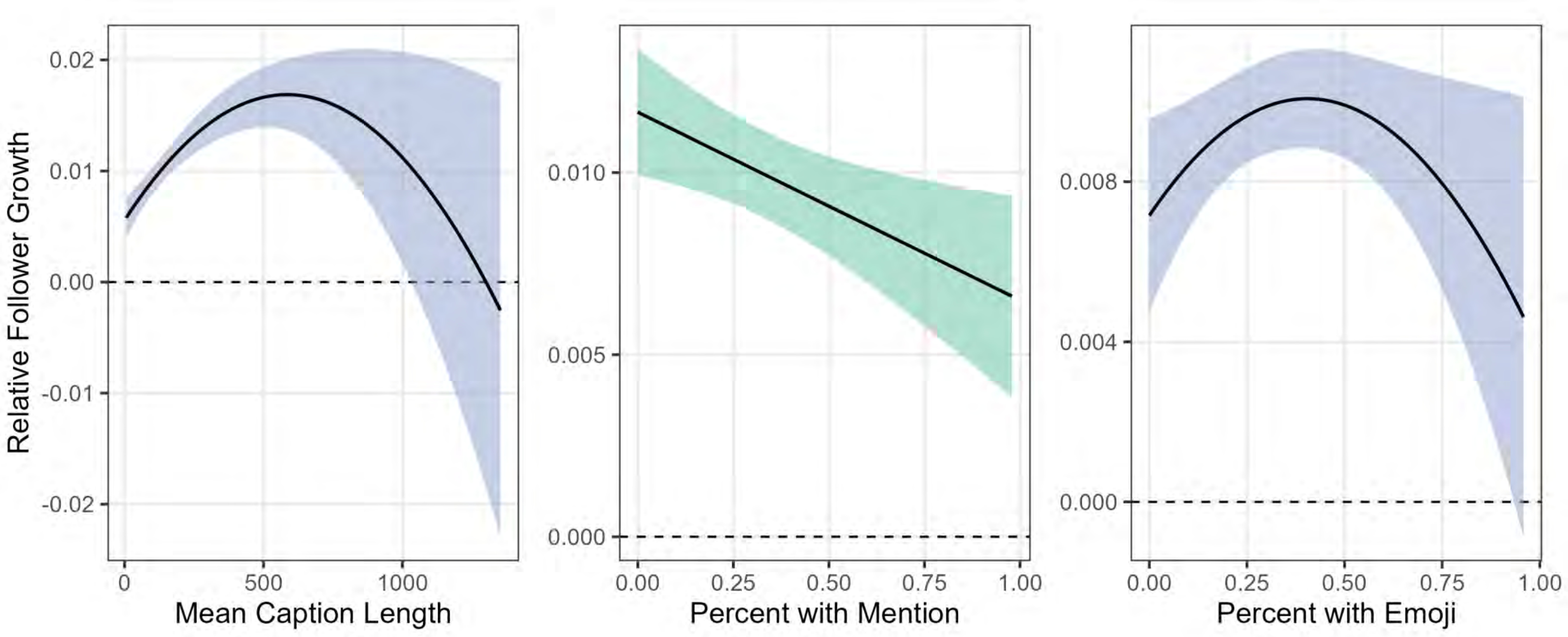
Predictors:

- Number of uploads per month
- Percent stories vs. posts / images vs. videos
- Percent explicit ads vs. non-ads
- Percent with captions / mean caption length
- Percent captions with mentions / hashtags
- Percent captions with emojis
- Percent images with one or more faces
- Mean smile intensity in images with faces
- Mean raciness (sexual content) in images

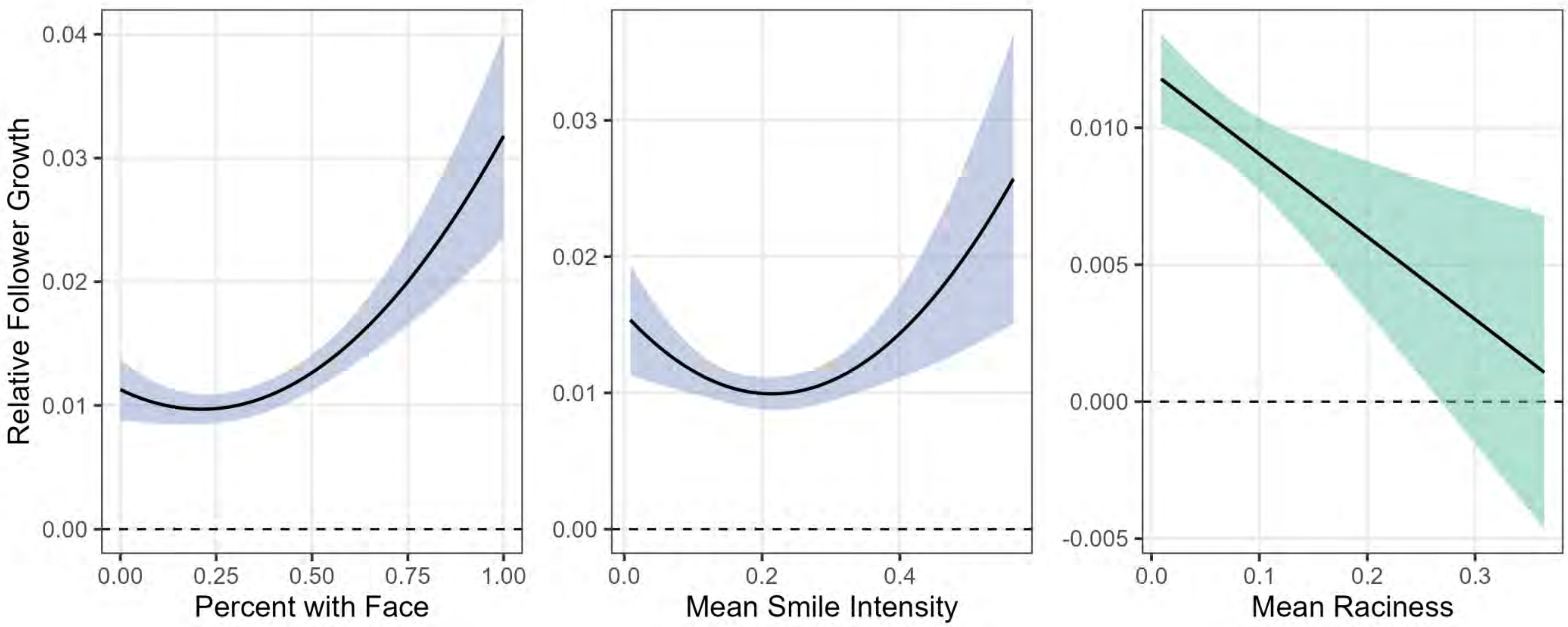
RESULTS (POSTING PATTERNS)



RESULTS (CAPTIONING)



RESULTS (PERSONAGE)



CONCLUSIONS

- The influencers with the highest relative growth in followers…
  1. Uploaded around 500 pieces of content per month (around 17 per day)
  2. Uploaded far more posts (permanent) than stories (disappear after one day)
  3. Uploaded far fewer pieces of content labeled as *explicit* advertisements
  4. Wrote text captions around 600 characters long (including links, mentions, tags)
  5. Included at-mentions (e.g., to other users) in fewer of their captions
  6. Included emojis in around 30–40% of their captions
  7. Uploaded images containing one or more human faces
  8. Uploaded images containing stronger smiles
  9. Uploaded images that were less “racy” or sexual

COVID AND SMILING RHYTHMS | 2

INTRODUCTION

- Smiling is a salient, common, and impactful socio-affective signal
- Photos posted to social media are a rich source of data for studies of smiling (large, frequent, global)
- Social behavior and affect are known to have temporal rhythms (e.g., daily, weekly, and seasonal)
- We planned to analyze temporal rhythms of smiling on Instagram
- *Then something happened in 2020…*

INTRODUCTION

- The COVID-19 pandemic was highly disruptive to many aspects of life
- Fear, uncertainty, loneliness, and loss were widespread negative emotions
- Social distancing and face masks changed social communication
- Lockdowns and work-from-home policies altered temporal rhythms
- We measured smiling on social media to study temporal rhythms before and during the pandemic

HYPOTHESES

HYPOTHESIS 1

At baseline, smiling will be *higher during weekend* days and show a seasonal cycle that *peaks during summer* months

HYPOTHESIS 2

Smiling will *decrease* during COVID's first year and then *partially return to baseline* during COVID's second year

HYPOTHESIS 3

COVID's first year will show a *dampened weekend* effect and *partially return to baseline* during COVID's second year

HYPOTHESIS 4

COVID's first year will show a *dampened seasonal* amplitude and *partially return to baseline* during COVID's second year

SOURCE & COUNTS

Partnered with **Whalar** (an international influencer management company)

- 1,905,424 images publicly uploaded
- 5,469 influencers on Instagram
  - 77.3% female, 21.2% male, 1.4% other
  - Age 18-64 (M=29.34, SD = 5.98)
- 76 countries of origin for influencers
  - 48.5% USA, 26.5% UK, 25% other
- 921 days from May 2019 – Oct 2021
  - All data were missing during Apr 2020

MEASURES

- Smile intensity was estimated using the OpenFace 2.0 toolkit (CV + ML system)

- Validated by 3 crowd-workers rating smile intensity (sample of 595 images)

$r = 0.41, 95\%$ CI: $[.34, .47]$

COMPARING TEMPORAL RHYTHMS

SEASONAL PERIODIC EFFECTS

$$\sin\left(t \times \frac{2\pi}{365}\right) \qquad \cos\left(t \times \frac{2\pi}{365}\right)$$

Amplitude

How large is the peak of the seasonal cycle?

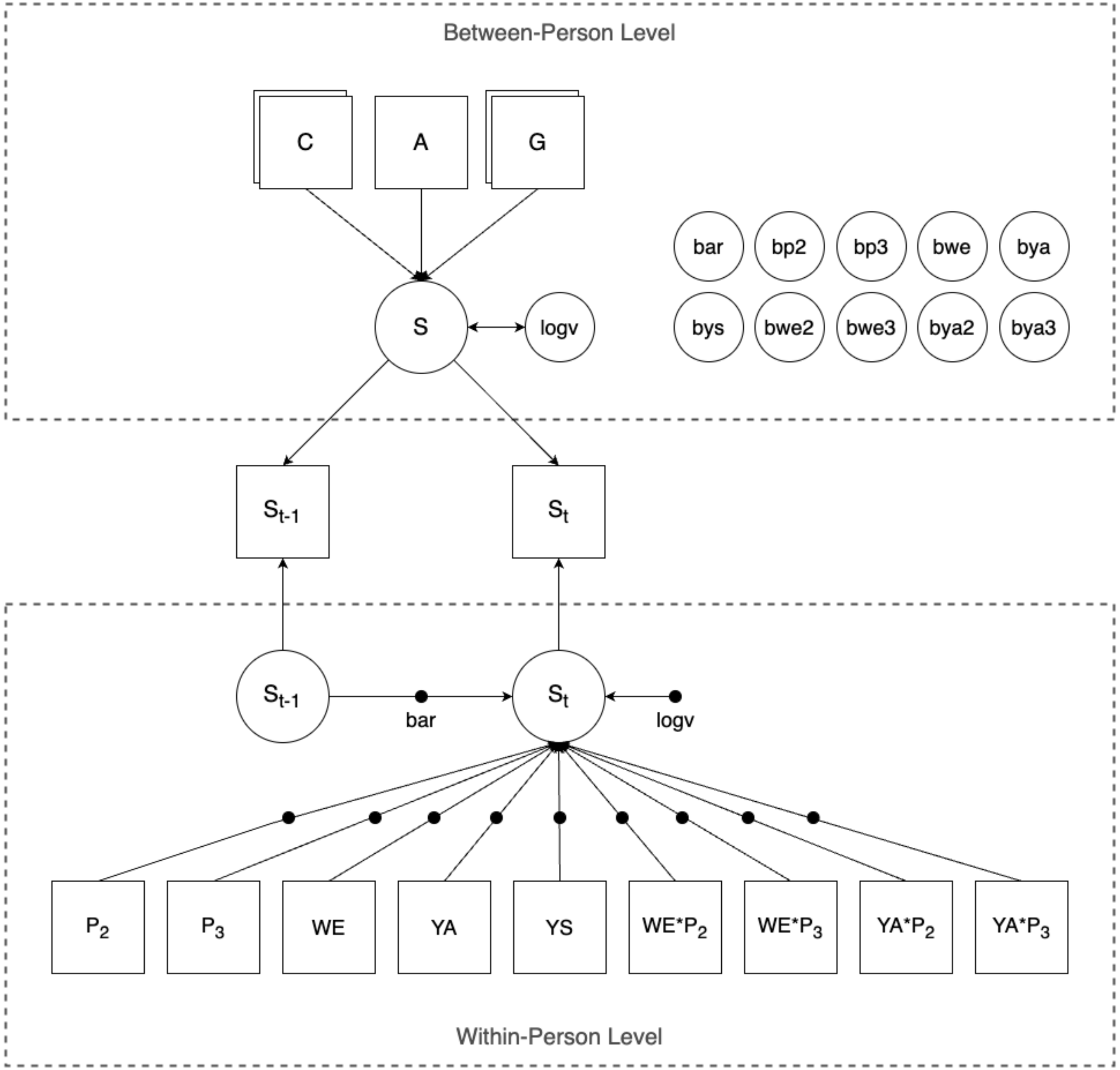
Phase Shift

When (in the year) does the cycle start?

$$A = \sqrt{\beta_{sin}^2 + \beta_{cos}^2} \qquad \phi = \text{atan2}(\beta_{sin}, \beta_{cos})$$

- Add seasonal periodic effect parameters
- Add a dummy code for weekend day
- Add dummy codes for study period (Baseline, COVID Year 1, COVID Year 2)
- Add interactions with period dummy codes
- *Does the weekend effect differ by period?*
- *Does seasonal amplitude differ by period?*
- *Does seasonal phase shift differ by period?*

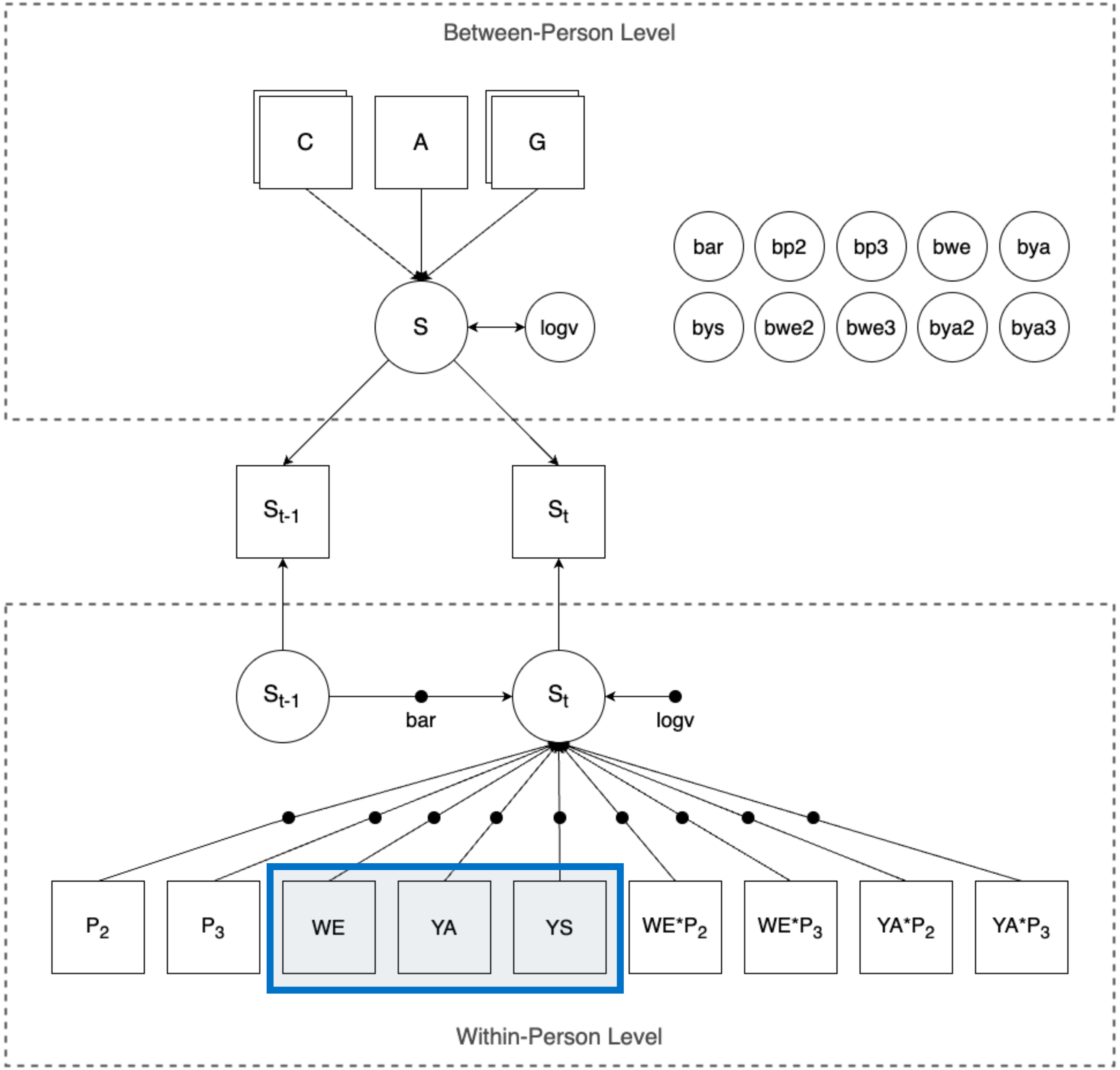
MODERATION BY PERIOD



DSEM PATH DIAGRAM

Dynamic Structural Equation Modeling

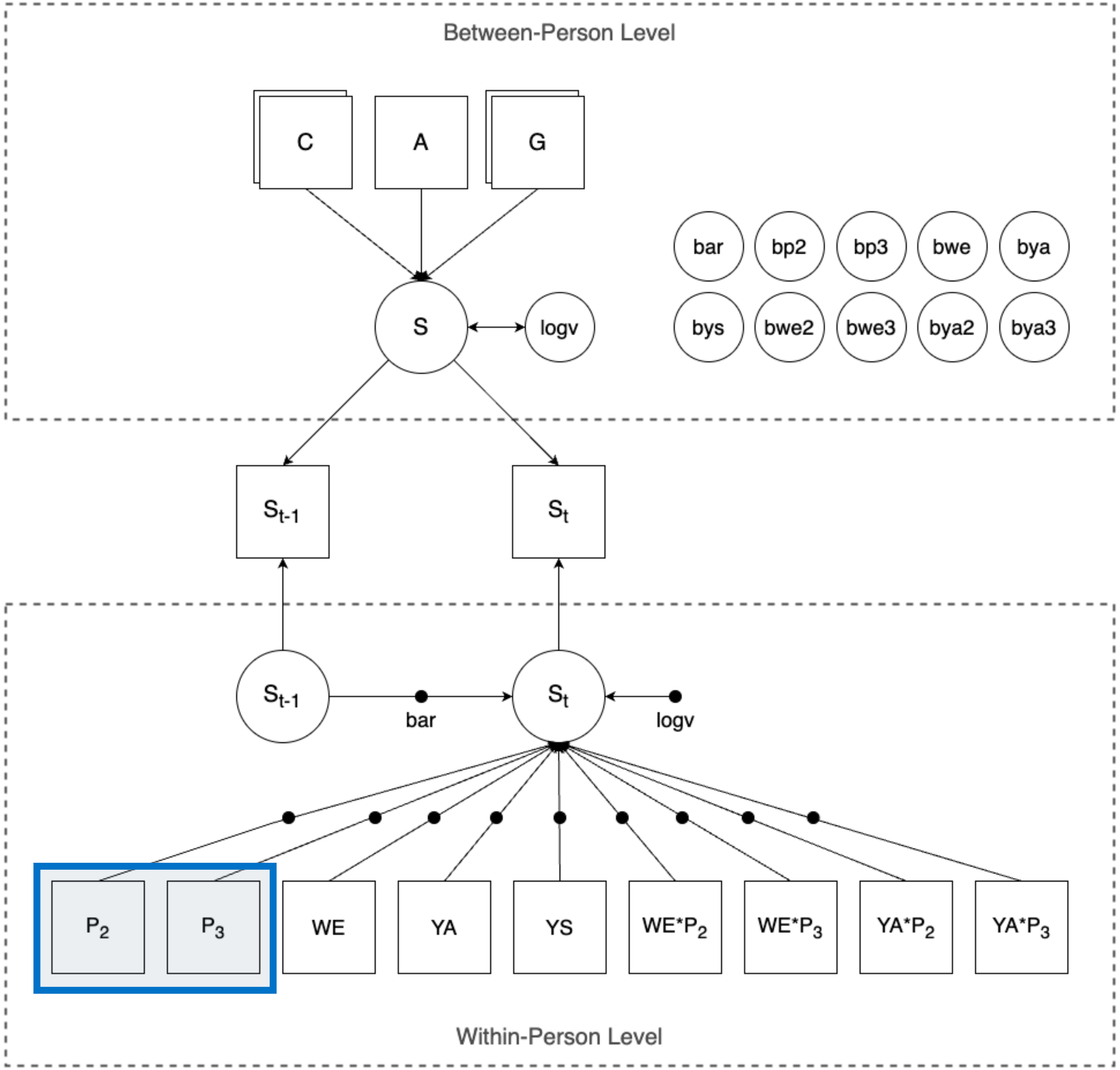
- Decompose into within/between levels
- Latent autoregressive/lagged effect
- Random intercepts, slopes, and errors
- Control for country, age, and gender



DSEM PATH DIAGRAM

Hypothesis 1

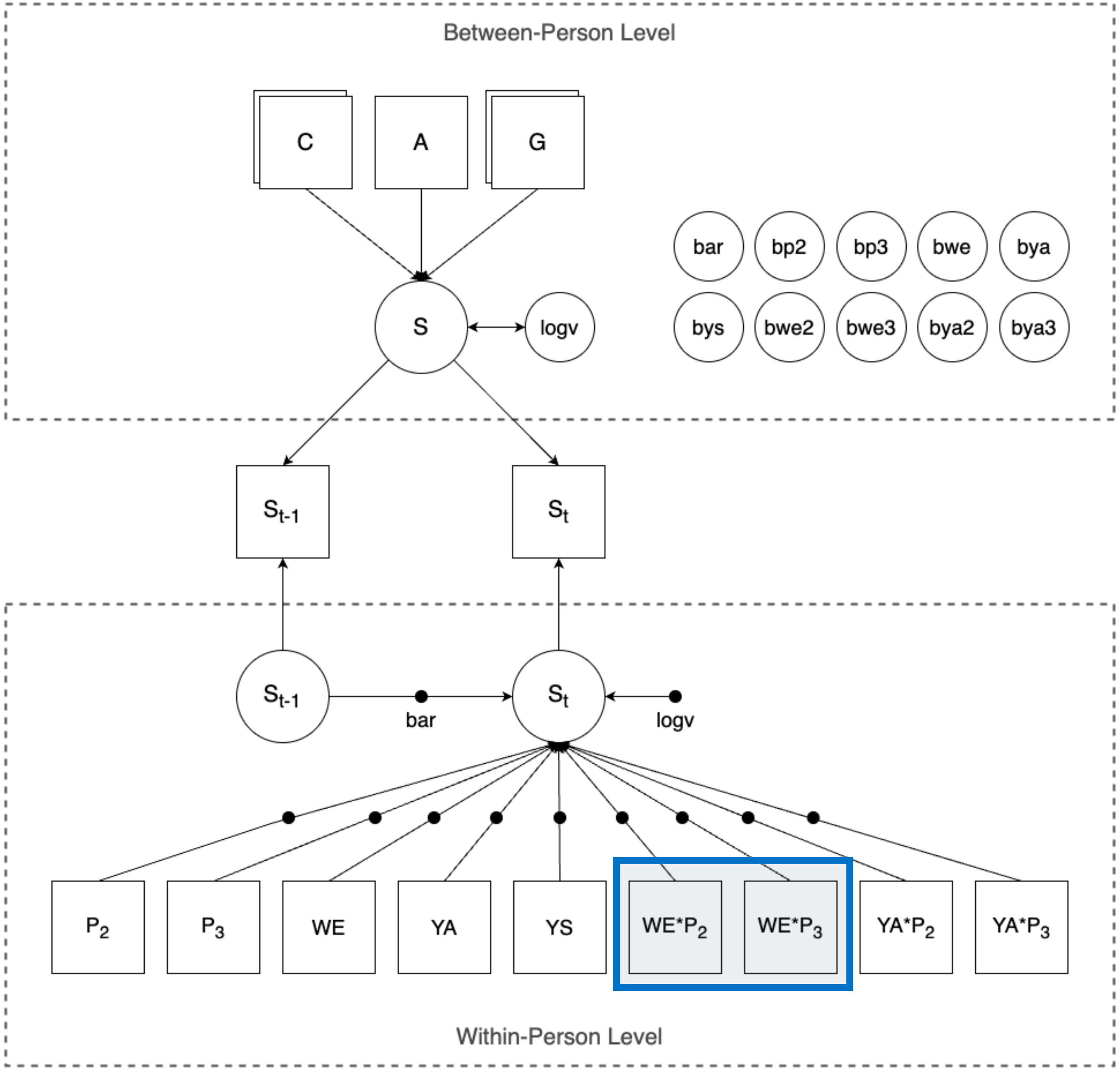
- *During the **baseline** period, did smiling show a **weekend** effect?*
- *During the **baseline** period, did smiling show a **seasonal** cycle?*



DSEM PATH DIAGRAM

Hypothesis 2

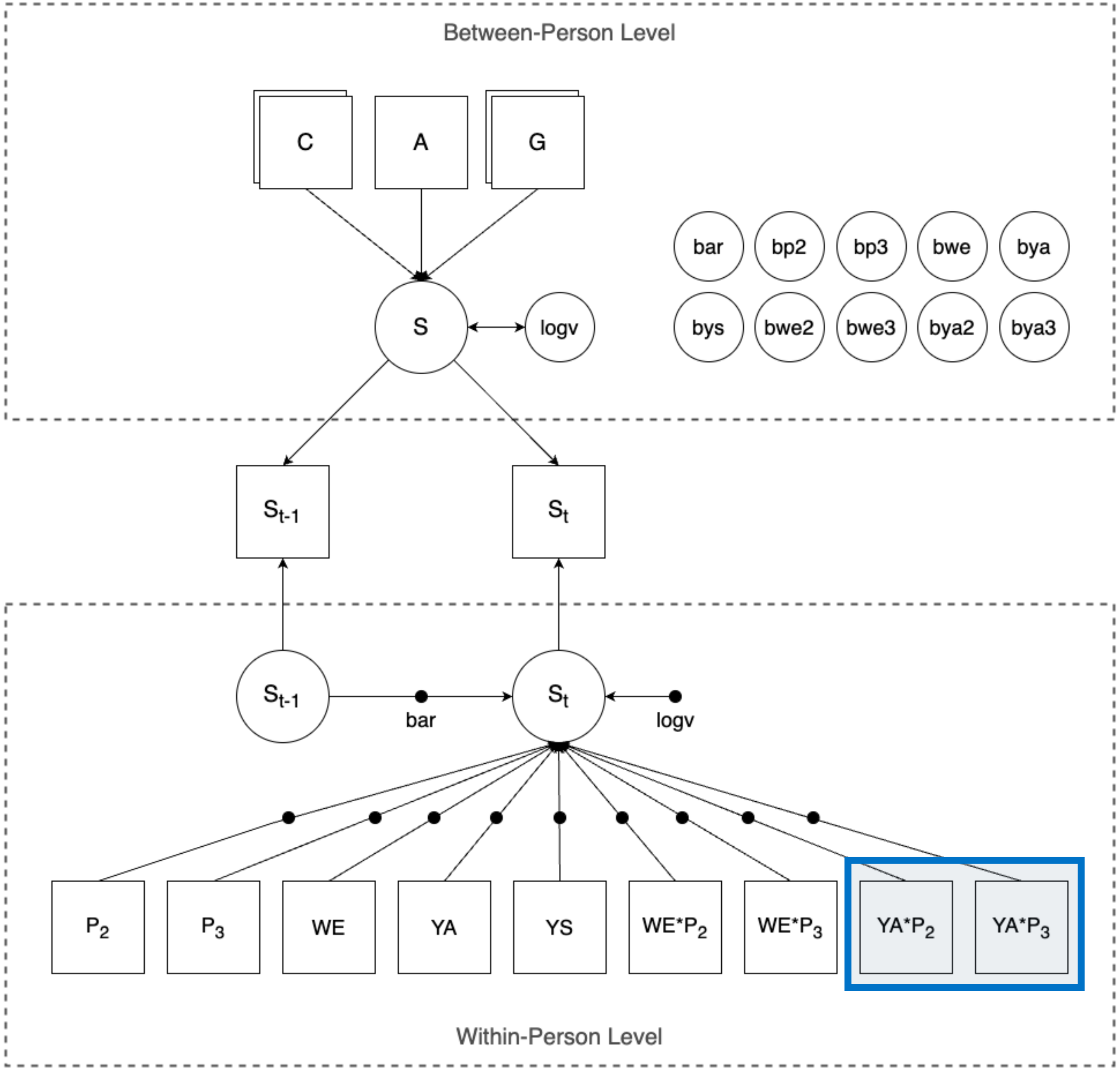
- *Did **average** smile intensity change from baseline to COVID **year one**?*
- *Did **average** smile intensity change from baseline to COVID **year two**?*



DSEM PATH DIAGRAM

Hypothesis 3

- *Did the **weekend** effect change from baseline to COVID **year one**?*
- *Did the **weekend** effect change from baseline to COVID **year two**?*



DSEM PATH DIAGRAM

Hypothesis 4

- *Did the **seasonal** effect change from baseline to COVID **year one**?*
- *Did the **seasonal** effect change from baseline to COVID **year two**?*

FIXED EFFECTS

Parameter	Est.	p	Sig.
Intercept	20.65	<.001	***
Age	0.73	<.001	***
Sex: Male	−4.03	<.001	***
Sex: Other	−2.11	<.001	***
Autoregression	0.03	<.001	***
Period 2	−0.11	.038	*
Period 3	0.32	<.001	***
Weekend	0.75	<.001	***
Yearly Amplitude	0.33	<.001	***
Yearly Phase Shift	0.00	.456	

	Est	p	Sig.
Weekend × Period 2	−0.14	.027	*
Weekend × Period 3	0.25	<.001	***
Amplitude × Period 2	−0.02	.400	
Amplitude × Period 3	0.52	<.001	***

FIXED EFFECTS

Parameter	Est.	p	Sig.
Intercept	20.65	<.001	***
Age	0.73	<.001	***
Sex: Male	−4.03	<.001	***
Sex: Other	−2.11	<.001	***
Autoregression	0.03	<.001	***
Period 2	−0.11	.038	*
Period 3	0.32	<.001	***
Weekend	0.75	<.001	***
Yearly Amplitude	0.33	<.001	***
Yearly Phase Shift	0.00	.456	

H1

	Est	p	Sig.
Weekend × Period 2	−0.14	.027	*
Weekend × Period 3	0.25	<.001	***
Amplitude × Period 2	−0.02	.400	
Amplitude × Period 3	0.52	<.001	***

FIXED EFFECTS

Parameter	Est.	p	Sig.
Intercept	20.65	<.001	***
Age	0.73	<.001	***
Sex: Male	−4.03	<.001	***
Sex: Other	−2.11	<.001	***
Autoregression	0.03	<.001	***
Period 2	−0.11	.038	*
Period 3	0.32	<.001	***
Weekend	0.75	<.001	***
Yearly Amplitude	0.33	<.001	***
Yearly Phase Shift	0.00	.456	

H2

	Est	p	Sig.
Weekend × Period 2	−0.14	.027	*
Weekend × Period 3	0.25	<.001	***
Amplitude × Period 2	−0.02	.400	
Amplitude × Period 3	0.52	<.001	***

FIXED EFFECTS

Parameter	Est.	p	Sig.
Intercept	20.65	<.001	***
Age	0.73	<.001	***
Sex: Male	−4.03	<.001	***
Sex: Other	−2.11	<.001	***
Autoregression	0.03	<.001	***
Period 2	−0.11	.038	*
Period 3	0.32	<.001	***
Weekend	0.75	<.001	***
Yearly Amplitude	0.33	<.001	***
Yearly Phase Shift	0.00	.456	

	Est	p	Sig.
Weekend × Period 2	−0.14	.027	*
Weekend × Period 3	0.25	<.001	***
Amplitude × Period 2	−0.02	.400	
Amplitude × Period 3	0.52	<.001	***

H3

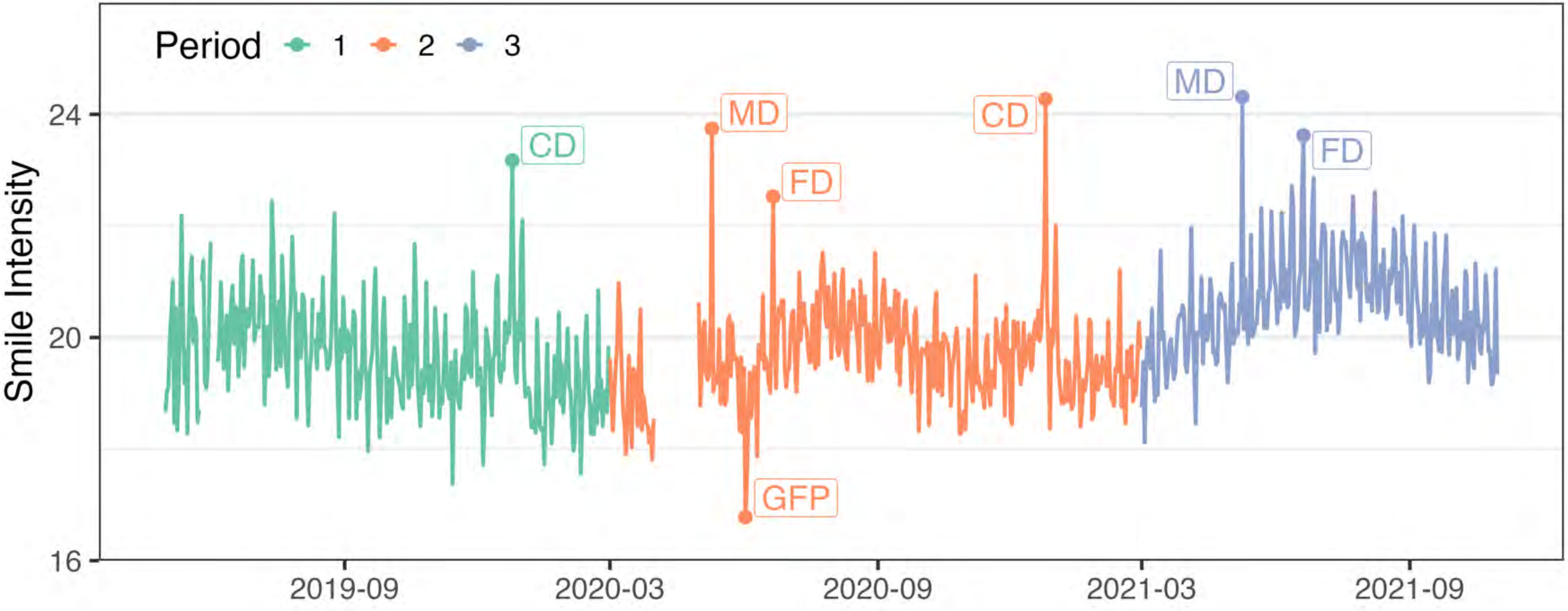
FIXED EFFECTS

Parameter	Est.	p	Sig.
Intercept	20.65	<.001	***
Age	0.73	<.001	***
Sex: Male	−4.03	<.001	***
Sex: Other	−2.11	<.001	***
Autoregression	0.03	<.001	***
Period 2	−0.11	.038	*
Period 3	0.32	<.001	***
Weekend	0.75	<.001	***
Yearly Amplitude	0.33	<.001	***
Yearly Phase Shift	0.00	.456	

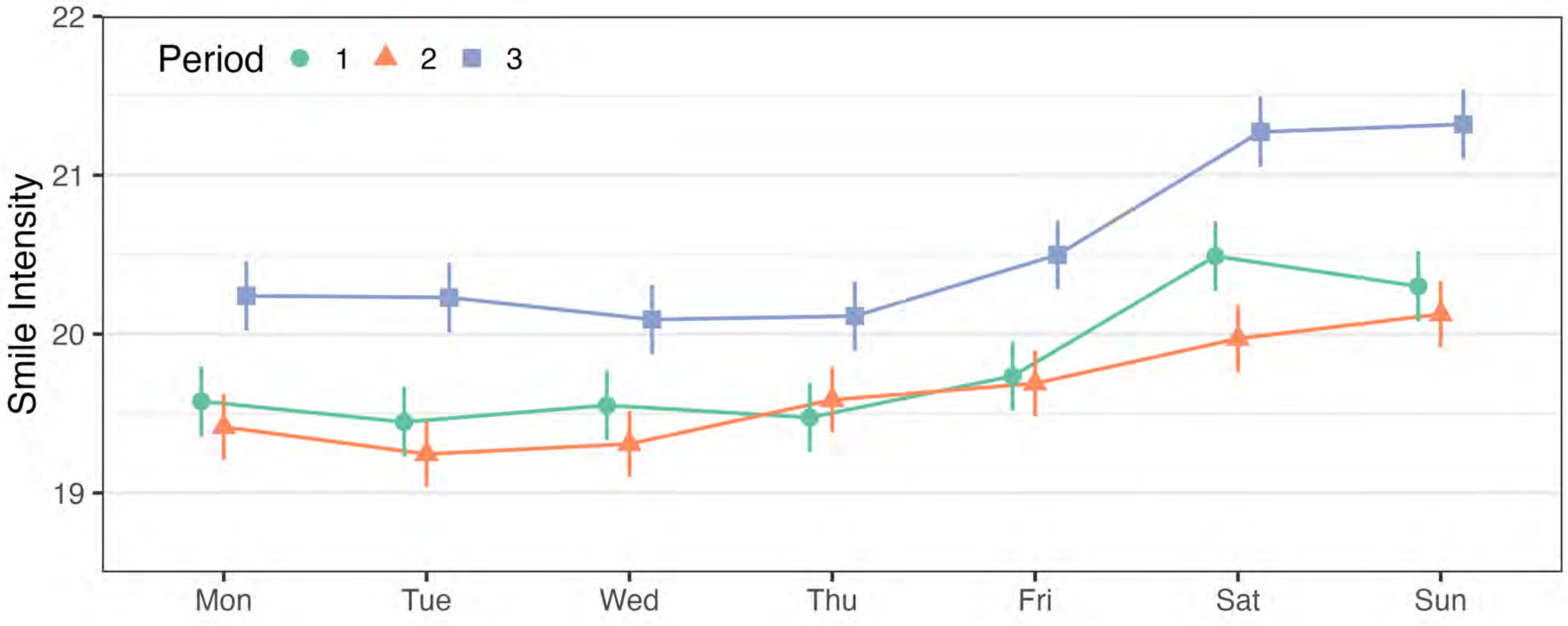
	Est	p	Sig.
Weekend × Period 2	−0.14	.027	*
Weekend × Period 3	0.25	<.001	***
Amplitude × Period 2	−0.02	.400	
Amplitude × Period 3	0.52	<.001	***

H4

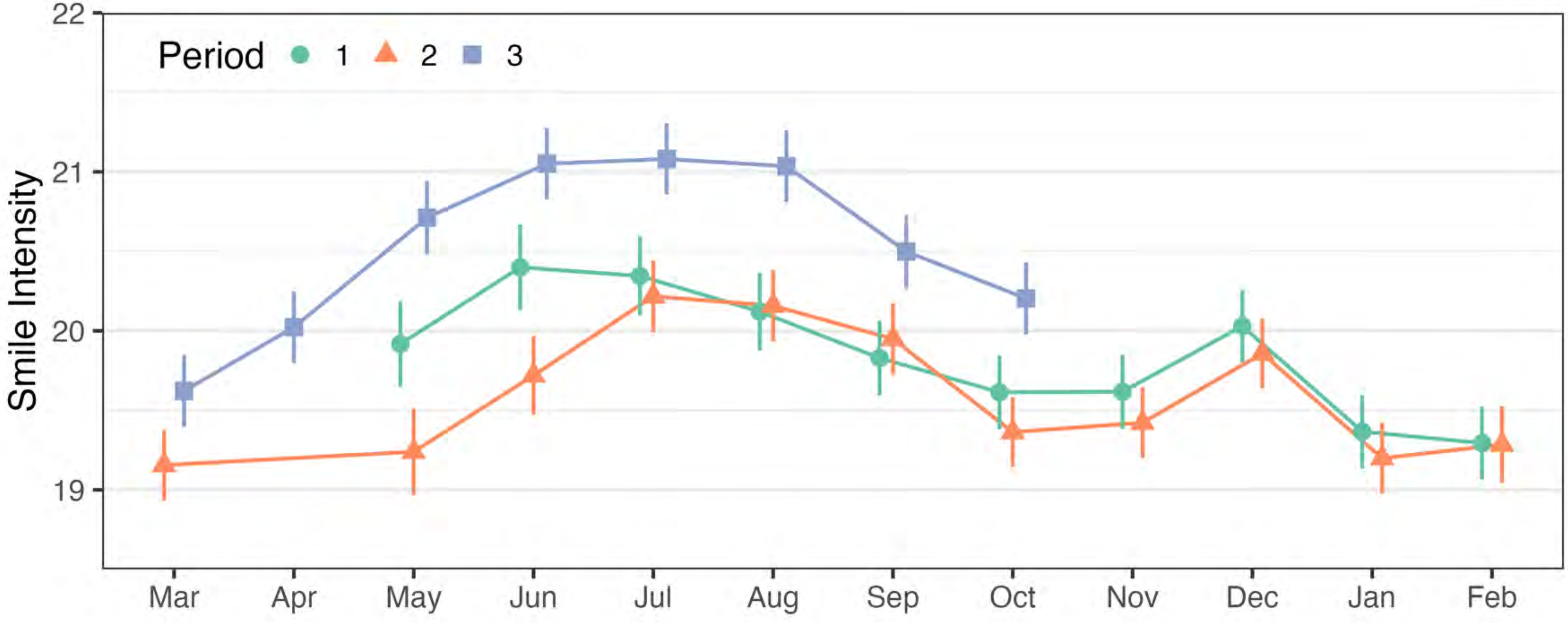
DAILY AVERAGES ACROSS PERIODS



WEEKDAY AVERAGES BY PERIOD



MONTH AVERAGE BY PERIOD



CONCLUSIONS

- The baseline (pre-COVID) year showed *weekend* and *seasonal* effects on social media smiling
- COVID year 1 showed *lower smiling* and a *dampened weekend* effect
- COVID year 2 showed *higher smiling,* an *amplified weekend* effect, and an *amplified seasonal* effect
- These results are consistent with a "rebound" effect as lockdowns ended
- Re-engagement with the environment and stronger influence of its properties

RESEARCH TEAM

DASHA YERMOL

PhD Student in Psychology
at University of Kansas

DANIEL MCDUFF

Staff Research Scientist
at Google

COLIN CAMPBELL

Assoc. Professor of Marketing at
University of San Diego

SARA ROSENGREN

Professor of Marketing & Strategy
Stockholm School of Economics

THANK YOU

JMGIRARD@KU.EDU

https://affcom.ku.edu

Affective Communication and Computing Lab

Departments of Psychology & Data Science