Supplemental Appendix to accompany Rights, Preacher, & Cole (*The danger of conflating level-specific effects of control variables when primary interest lies in level-2 effects*):

Empirical example predicting cognitive response using each of the four modeling approaches

To illustrate the potential distortion of results that can occur with the *conflated-x model*, we present an empirical example wherein substantive interest is in the effect of a level-2 variable after controlling for a level-1 variable. Data were obtained via an adaptation of Cole et al.'s (2014) Behind Your Back (BYB) procedure. In this procedure, 272 adolescents listened to recordings of 21 conversations of a boy and girl talking about an absent third person, with the gender of the third person matching that of the participant. The content of the conversations ranged from mild to mean. Participants were given the following instructions: "*Try to imagine that you hear two people talking about you behind your back. Below is what they say. Read each conversation as you listen to them and pretend you could actually hear them. Also pretend that they are talking about you. After each conversation, there will be some questions for you to answer. Circle a number to answer each question*." Participants were asked "*How mean were they being?*" and "*If they were talking about you, how sad would this make you feel?*" on 5-point Likert scales (1 = *not at all*, 2 = *just a little*, 3 = *somewhat*, 4 = *pretty much*, 5 = *a lot*). They were similarly asked scenario-specific questions regarding their cognitive responses to the scenario; e.g.: "*If you heard this, how much would it make you think: 1) I am an idiot. 2) I don't like myself.*" again on a 5-point scale. Data thus consisted of 21 conversation-specific observations nested within each adolescent.

Of substantive interest is the effect of an adolescent's propensity for *negative appraisal* (*NA*) of events (level-2 predictor) on maladaptive *cognitive reaction* (*COG*) to events, above and beyond that attributable to one's event-specific *emotional reaction* of sadness (*SAD*; level-1 control variable). In other words: do adolescents who tend to interpret social events more negatively tend to, in turn, have more negative cognitive reactions to said events, regardless of their emotional response? Prior research has linked both negative appraisals and emotional response to cognitive reactions that contribute to depressive schema (e.g., Davidson, 1998; Everaert, Koster, & Derakshan, 2012; Platt, Waters, Schulte-Koerne, Engelmann, & Salemink, 2017); however, studies have rarely assessed both negative appraisals and emotional reactions simultaneously, making unclear the contribution of each individually. In this example, we wish to assess the impact of one's propensity for negative appraisals on cognitive reactions to events, controlling for the emotional response of sadness, in order to isolate the purely cognitive component of depressive schema development as distinct from the affective component. It could be the case, for instance, that the apparent impact of negative appraisal can be explained solely through its association with emotional reaction, or conversely that negative appraisals may exert an important influence on cognitive reactions above and beyond its relation with emotional reactions.

We fit each of the four models in manuscript Figure 1 to these data. The *fully-disaggregated-x* model, is given as:

$$\begin{aligned} COG_{ij} &= \gamma_{00} + \gamma_{x_w}(SAD_{ij} - SAD_{\bullet j}) + \gamma_{x_b} SAD_{\bullet j} + \gamma_z NA_j + u_{0j} + e_{ij} \\ e_{ij} &\sim N(0, \sigma^2) \\ u_{0j} &\sim N(0, \tau_{00}) \end{aligned} \quad \text{(S1)}$$

NA, a level-2 variable, is the person-mean of subjective meanness (described above) across all 21 conversations; *SAD* is the person's rating of sadness in response to the conversation; and *COG* is the person's cognitive evaluation rating (mean of two responses).

These results are displayed in Supplemental Figure 1 Panel A. Here both the within-person and between-person slopes of *SAD* are positive and significant, that is, adolescents who tend to report higher levels of sadness tend to have more negative reactions (slope of $SAD_{\bullet j}$) to conversations, and when adolescents report more sadness relative to their baseline level, they tend to have more negative reactions (slope of $SAD_{ij} - SAD_{\bullet j}$) to that particular conversation. However, the slope of *NA*, which is of primary interest, is small and nonsignificant. These results suggest that the association between *NA* and *COG* may be attributed instead to the former's association with *SAD*.

In interpreting these results, it is intuitive to think that adolescents who are more likely to give negative appraisals (higher levels of *NA*) are, in turn, more likely to have greater emotional response to events (higher levels of *SAD*). Results here suggest that it may be the emotional response that is causing the cognitive reaction, rather than the propensity for negative appraisal itself; this is consistent with results found in Cole et al. (2019). Although the exact mechanism can be further considered and explored, the primary point we make here is that there seems to be no or little impact of *NA* when appropriately controlling for *SAD*.

As an alternative approach, to effectively control for *SAD* with respect to level-2 predictors, we can simply include the person-mean of *SAD* as control variable, that is, fit the *x-mean model*:

$$
\begin{aligned}
COG_{ij} &= \gamma_{00} + \gamma_{x_b} SAD_{\bullet j} + \gamma_z NA_j + u_{0j} + e_{ij} \\
e_{ij} &\sim N(0, \sigma^2) \\
u_{0j} &\sim N(0, \tau_{00})
\end{aligned} \qquad \text{(S2)}
$$

As expected, we get the same estimate for the slope of *NA* as we did for the *fully-disaggregated-x model*, as shown in Supplemental Figure 1 Panel B. We do not obtain information about potential within-person effects of *SAD* in this model as we do in the *fully-disaggregated-x model*, but this is of little consequence if focus is strictly on the effect of NA.

Now suppose instead we were to fit the *no-x model*:

$$
\begin{aligned}
COG_{ij} &= \gamma_{00} + \gamma_z NA_j + u_{0j} + e_{ij} \\
e_{ij} &\sim N(0, \sigma^2) \\
u_{0j} &\sim N(0, \tau_{00})
\end{aligned} \qquad \text{(S3)}
$$

Results for this model are displayed in Supplemental Figure 1 Panel C. This model would suggest that there is, in fact, an impact of *NA* on *COG*. Thus, when we fail to control for *SAD*, we might erroneously assume that *NA* has a direct effect on *COG*.

Now suppose that we had followed current recommendations and simply grand-mean-centered *SAD* as a control variable, that is, fit the *conflated-x model*:

$$
\begin{aligned}
COG_{ij} &= \gamma_{00} + \gamma_{x_c} (SAD_{ij} - SAD_{\bullet\bullet}) + \gamma_z NA_j + u_{0j} + e_{ij} \\
e_{ij} &\sim N(0, \sigma^2) \\
u_{0j} &\sim N(0, \tau_{00})
\end{aligned} \qquad \text{(S4)}
$$

Results for this model are displayed in Supplemental Figure 1 Panel D. Similar to the *no-x model*, this would suggest that there is an effect of *NA* on *COG*, even when controlling for *SAD*. However, this model does not control for *SAD appropriately* because it does not disaggregate the within-person and between-person effects. As we can see by looking at the point estimates in Supplemental Figure 1, the conflated effect of SAD (0.36) in Panel C is weighted heavily towards the estimated within-effect (0.35) shown in Panel A, which is substantially smaller than the estimated between-effect (0.69). We know from the bias derivation (see manuscript Equation 17) that this difference in the between-effect and the conflated effect can drive the bias in the slope of the level-2 variable of interest. In this case, this difference yields a slope estimate for *NA* for the *conflated-x model* of .34.

Summary. Clearly, by failing to control for the level-1 variable appropriately, we get very different results than if we had done so. Controlling inappropriately with the *conflated-x model* would suggest that one's propensity for negative appraisals has a direct impact on cognitive response to events that is comparable to the direct impact of emotional reaction. Controlling appropriately with either the *fully-disaggregated-x* model or the *x-mean* model suggests that the association of negative appraisals and cognitive response might be better understood through the former's relation with emotional reaction. We use this example to stress that using conflated slopes of level-1 variables as a control for level-2 variables is not appropriate and can severely distort results and profoundly alter the substantive implications.

References

Cole, D. A., Martin, N. C., Sterba, S. K., Sinclair-McBride, K., Roeder, K. M., Zelkowitz, R., & Bilsky, S. A. (2014). Peer victimization (and harsh parenting) as developmental correlates of cognitive reactivity, a diathesis for depression. *Journal of Abnormal Psychology, 123,* 336-349.

Cole, D. A., Zelkowitz, R. L., Nick, E. A., Lubarsky, S. R., & Rights, J. D. (2019). Simultaneously examining negative appraisals, emotion reactivity, and cognitive reactivity in relation to depressive symptoms in children. *Development and Psychopathology*, *31*, 1527-1540.

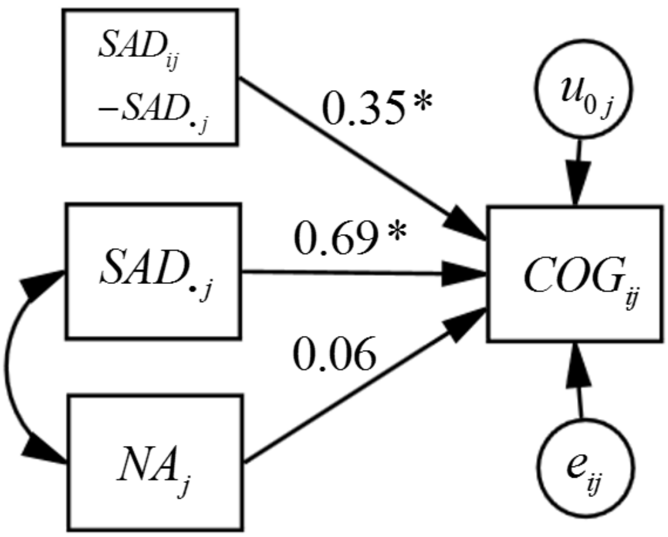
Davidson, R. J. (1998). Affective style and affective disorders: Perspectives from affective neuroscience. *Cognition and Emotion*, *12*, 307-330.

Everaert, J., Koster, E. H. W., & Derakshan, N. (2012). The combined cognitive bias hypothesis in depression. *Clinical Psychology Review*, *32*(5), 413-424.

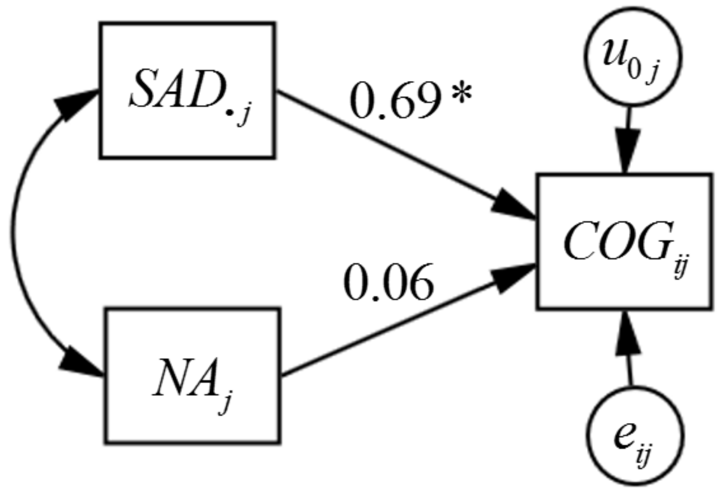
Platt, B., Waters, A. M., Schulte-Koerne, G., Engelmann, L., & Salemink, E. (2017). A review of cognitive biases in youth depression: Attention, interpretation and memory. *Cognition and Emotion*, *31*(3), 462-483.

Supplemental Figure 1. Empirical example results: Predicting cognitive response using each of the four modeling approaches

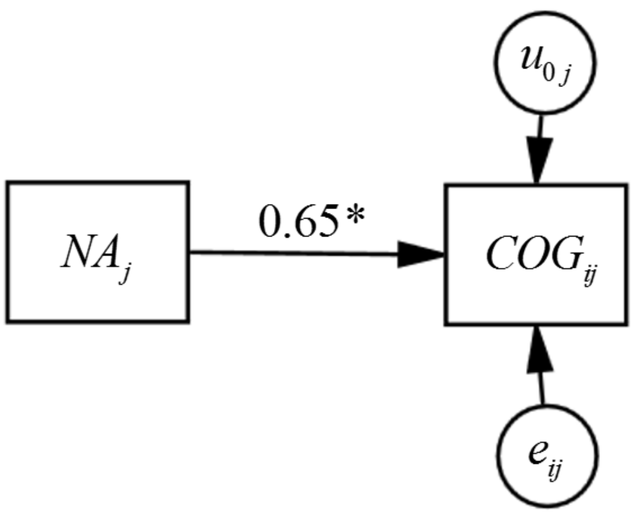
Panel A: *Fully-disaggregated-x model*



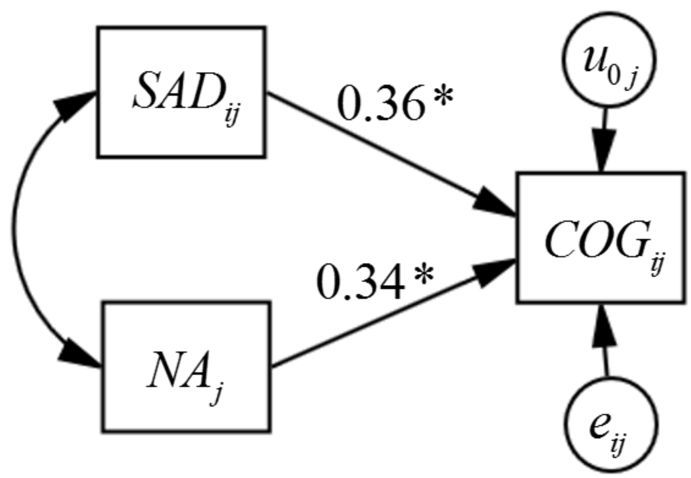
Panel B: *x-mean model*



Panel C: *No-x model*



Panel D: *Conflated-x model*



*p < .005