FINAL REPORT.

A. PRIMARY OBJECTIVES.

The maternal still-face (SF) is an established experimental infant social stressor that typically results in increased infant distress (Adamson & Frick, 2003; Mesman, van IJzendoorn, & Bakermans-Kranenburg, 2009; Tronick, Als, Adamson, Wise, & Brazelton, 1978). While this is generally true, past research from our own laboratory in addition to other laboratories suggests major individual differences in both infants' response to the SF and their subsequent recovery.

The primary goal of this project was to explore individual differences in infants' reactivity and recovery to the SF via changes in salivary stress biomarkers, specifically changes in salivary cortisol and alpha-amylase (sAA). To do so, infants and their mothers participated in a series of face-to-face play and maternal SF episodes until either 1) the infant became distressed (i.e., consistent crying for 30 seconds) or 2) the infant completed three SF-Play repetitions (see Figure 1). This unique design allowed for the calibration of each infant's stress response to the social stressor and thereby provided a more accurate way of assessing changes in behavioral and physiological systems in response to stress. To the authors' knowledge, this is the first study that explored changes in salivary stress biomarkers by tethering actual infant distress to the timing of post-stress samples.

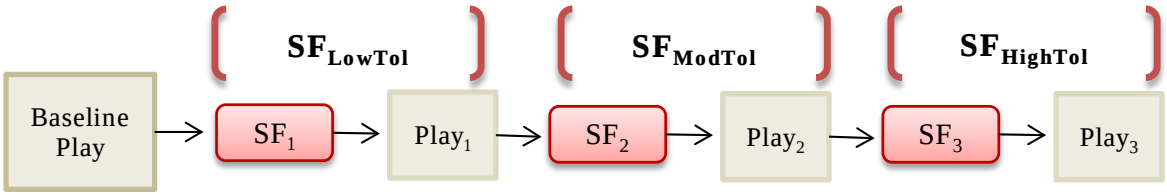

Figure 1. Infants and their mothers participated in a modified FFSF paradigm. The Still-Face(SF)-Play episodes continued until the infant became distressed or for a maximum of three repetitions. Infants' stress tolearnce (low, moderate, high) was determined by the timing of their distress.

B. PARTICIPANTS.

Thirty-two mother-infant (13 female) dyads participated in the study. Two dyads were excluded entirely from the study. One dyad did not complete the baseline play

episode so the testing session was stopped before the first SF stressor. For the other dyad the research assistant (RA) terminated the testing session before the infant cried for 30 seconds and therefore violated study criteria. Thirty dyads remained in the final sample.

C. CODING AND DATA PREPARATION.

Behavioral Coding. Dropout groups were determined in real-time during testing sessions. Dyads were classified into dropout groups based upon when, or if, the infant became visibly upset (i.e., cried for 30 consecutive seconds). Videos were then analyzed offline to ensure reliability. Two coders, blind to hypotheses, coded infant negative affect (negative affect, cry, other) using 1-second time intervals (Mangold Interact software, Mangold International, GmbH (Ed.)). Durations of continuous crying were checked to ensure that all dropout infants cried for at least 30 seconds (+/- 5 seconds). Thirty mother-infant dyads (13 female) were included in the behavior analyses. Of the 30 dyads, 10% (n=3) dropped out of the paradigm following the first SF episode, 23% (n=7) after the second SF episode, and 17% (n=5) after the third SF episode. The remaining 50% (n=15) completed the entire paradigm. Given the distribution of participants, dropout categories were collapsed to create two main stress groups, dyads who dropped out during the paradigm ($Group_{Drop}$ n=15) and those who completed the paradigm ($Group_{NoDrop}$ n=15).

Salivary Samples. Using a pre/post-stress design, 5 saliva samples were collected from infants to observe changes in salivary cortisol and salivary alpha-amylase (sAA). Saliva samples were collected with Sorbettes (Salimetrics, State College, PA), specifically designed for infants and small children. Two baseline samples were taken 8 minutes apart before the start of the testing paradigm followed by three post-paradigm samples taken at 5-, 20- and 40 minutes. The second baseline sample was used in the analyses. The 5-minute and 20-minute post-paradigm samples served as the reactivity and recovery samples for sAA while the 20- and 40-minute samples represented the reactivity and recovery samples for cortisol. Saliva samples were immediately stored in a -20 C degree freezer before being moved to a -80 C degree freezer for longer-term storage until shipped overnight on dry-ice for assay at Dr. Douglas Granger's Institute for Interdisciplinary Salivary Bioscience (Arizona State University). Samples were thawed and centrifuged (to remove mucins) and assayed in duplicate for salivary cortisol using a commercially available immunoassay specifically designed for use with saliva without modification to the manufacturer's (Salimetrics, Carlsbad, CA) recommended protocol. The test volume was 25 μl, the range of sensitivity was from .007 to 3.0 μg/dL, and intra- and inter-assay coefficients of variation were on average less than 10% and 15% respectively.

Data Preparation. One infant was dropped from the final sample due to taking a medication thought to affect salivary cortisol (Granger, Hibel, Fortunato, & Kapelewski, 2009). Individual saliva samples were excluded for a number of methodological reasons including missing samples (due to extreme infant distress), insufficient sample volume, and assay interference. Samples were also excluded if infants were fed within 10 minutes of taking a sample. While rinsing infants' mouths with water should reduce any potential effects of milk or formula, doing so in close proximity to the sample time may dilute cortisol concentrations, as per assay manufacturer's guidelines (Salimetrics, LLC). The baseline sample and one post-stress sample (5-minute for sAA and either 20- or 40-minute for cortisol) were required from all participants. After excluding individual saliva samples, an additional three infants (n=2 no baseline cortisol sample, n=1 no post-paradigm sample) were dropped from the final cortisol and sAA analyses.

All data were explored within groups for possible univariate (> +/- 3 SD from the mean) and multivariate (Mahalanobis Distance, $p < 0.001$) outliers in addition to violations of normality across all dependent variables. Transformed data were used in analyses while raw data were used for descriptive statistics presented in figures and tables for clarity.

D. RESULTS.

Behavior. Duration of time was changed to proportion of time to account for differences in total episode times (e.g., episodes that were shortened due to more than 30 seconds of infant crying). Behavioral data was transformed using an arcsine transformation appropriate for proportion data (2*(arcsine(p1/2)) to correct normality issues (Howell, 2002). Thirty infants ($Group_{Drop}$ n = 15, $Group_{NoDrop}$ n =15) were included in the behavior analysis. In addition to excessive crying, $Group_{Drop}$ infants were also generally more negative ($M = 1.014$, $SD = 0.275$) throughout the entire testing session compared to $Group_{NoDrop}$ infants ($M = 0.333$, $SD = 0.278$), $t(28) = -6.741$, $p < 0.001$, $d=2.462$.

Cortisol. Raw cortisol values were transformed using a $\log_{10}$ transformation to address a significant positive skew. Only one outlier remained following this transformation and the participant was subsequently dropped. As a result, 25 infants ($Group_{Drop}$ n = 11, $Group_{NoDrop}$ n = 14) were included in the cortisol analyses.

When considering the effect of dropout group on the three cortisol time points (baseline, 20-minutes post, 40-minutes post), there were no main effects or interactions (see Table 1).

The lack of findings when considering both the 20 and 40 minute samples may have been due to differences in peak cortisol times. Some infants' cortisol peaked at 20 minutes while others' cortisol peaked at 40 minutes. When considering post-testing

	Wilks' λ	F(df)	$\eta^2_{partial}$
Group		0.004(1,23)	<0.001
Sample	0.923	0.923(2,22)	0.077
Group x Sample	0.959	0.470(2,22)	0.041

Table 1. Findings from the 2 (Group: Drop, No Drop) x 3 (Sample: Baseline, 20-minute, 40-minute) cortisol analyis. There were no significant main effects or interactions (*ps* > 0.1).

session peak times, there was no relationship between group membership and post-peak cortisol times. Fifty-five percent of Group$_{Drop}$ infants had a peak cortisol response 20 minutes post-paradigm and 45% peaked 40-minutes post-paradigm while 57% of Group$_{NoDrop}$ infants' peaked 20 minutes posttest and 43% peaked 40-minutes posttest ($p > 0.999$, Fischer's exact test). When collapsing across dropout group, there was also no difference in peak cortisol time, $X^2(1,N=25) = 0.16$, $p = 0.689$. For these reasons, the greatest posttest cortisol value was used as the peak value regardless of post-paradigm sample time.

There was a positive relationship between baseline cortisol and peak cortisol, $r(25) = 0.842$, $p < 0.001$, with greater average baseline cortisol values were associated with greater peak cortisol values. While both groups showed this positive relationship (Group$_{Drop}$ $r(11) = 0.877$, $p < 0.001$ and Group$_{NoDrop}$ $r(14)=0.764$, $p = 0.001$), there was no difference in relationship strength between the groups, $Z = 0.77$, $p = 0.441$.

Two dependent variables were used to explore changes in cortisol. First, infants' baseline cortisol and highest post-paradigm peak cortisol were used to calculate a **peak cortisol reactivity difference score** (i.e., greatest post – baseline). Unexpectedly, there was no difference in cortisol reactivity between groups, $t(23) = 0.152$, $p = 0.880$, $d=0.060$. As shown in Figure 2, Group$_{Drop}$ and Group$_{NoDrop}$ infants showed similar changes in cortisol. Second, categories representing **cortisol stress responses** were used. Infants were categorized as having a ***Stress Response*** if the difference between the greatest post-paradigm cortisol sample and the baseline sample was both positive and greater than 10% (Granger et al., 2012). If the difference was negative and/or less than 10%, the infant was classified as ***No Stress Response***. Overall infants were more likely to have a cortisol stress response (68%) than not, X^2 (N=25) = 3.24, $p = 0.072$ (trend). However, unexpectedly, there was no association between dropout group and infants' cortisol stress response. Equal numbers of Group$_{Drop}$ infants and Group$_{NoDrop}$ infants (73% respectively) had a stress response, $p > 0.999$, Fisher's exact test). Adding to this finding, there was also no relationship between cortisol stress response group and total negative behavior, $F(1,23) = 1.640$, $p = 0.213$, $\omega^2=0.0249$.

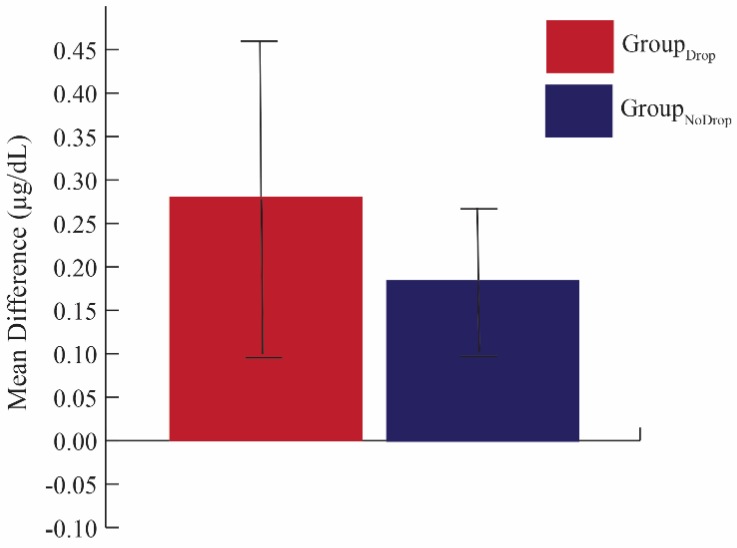

Figure 2. While Group$_{Drop}$ infants had a greater mean peak cortisol reactivity difference score compared to Group$_{NoDrop}$ infants, the mean difference was not significant ($p > 0.1$).

s*AA.* Raw sAA values were squared to address positive skew. One outlier remained following the transformation and was subsequently dropped from analyses. In all, 26 infants were included in the sAA analyses (Group$_{Drop}$ n = 13, Group$_{NoDrop}$ n=13).

When considering the three sAA time points (baseline, 5-minutes post, 20-minutes post), there was a main effect of sample time (see Table 2). Unexpectedly, baseline sAA was greater than both the 5-minute (reactivity) and the 20-minute (recovery) post-paradigm samples regardless of dropout group (see Figure 3).

Similar to cortisol, there was a positive relationship between infants' baseline sAA and the 5-minute post-paradigm sample, $r(26) = 0.884$, $p < 0.001$, where greater average BL sAA levels were associated with greater post-paradigm reactivity levels. Both groups showed this relationship (Group$_{Drop}$ $r(13) = 0.870$, $p < 0.001$ and Group$_{NoDrop}$ r(13)

	Wilks' λ	F(df)	$\eta^2_{partial}$
Group		0.383(1,23)	0.016
Sample	0.725	4.163(2,22)*	0.029
Group x Sample	0.975	0.278(2,22)	0.025

* significant at $p < 0.05$

Table 2. Findings from the 2 (Group: Drop, No Drop) x 3 (Sample: Baseline, 20-minute, 40-minute) sAA analyis. There was a significant main effect of sample time where baseline samples were greater than both the post-paradigm 5-minute and 20-minute samples for both groups.

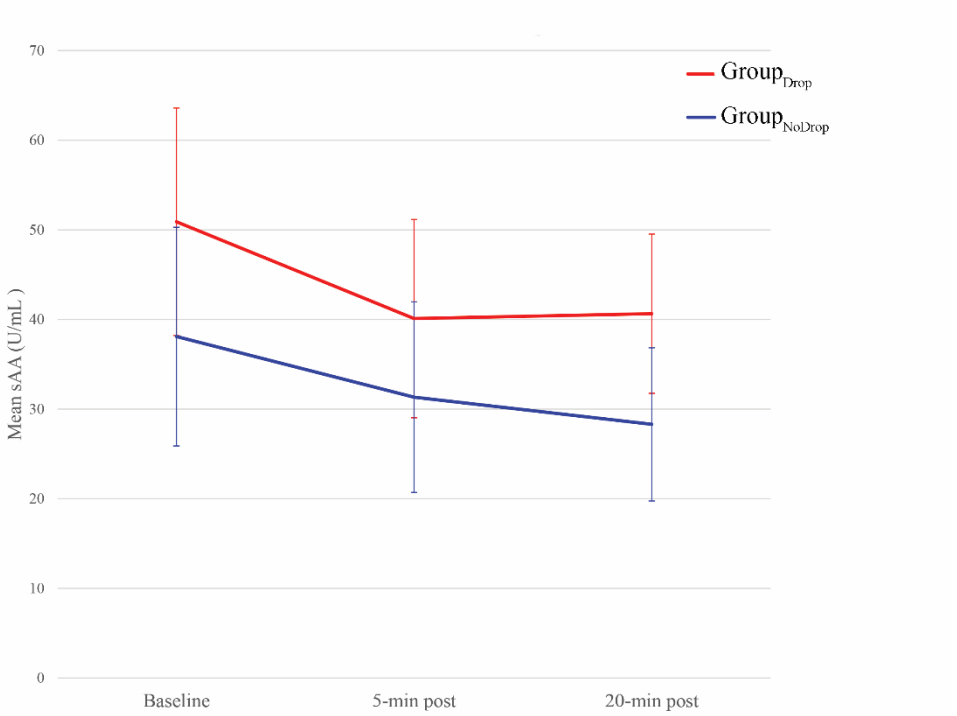

Figure 3. Overall infants' baseline sAA sample was significantly greater than the 5-minute and the 20-minute post-paradigm samples.

= 0.910, $p < 0.001$) and there was no difference in relationship strength between groups, Z=0.43, p= 0.668.

A reactivity difference score was calculated using both the average baseline sAA sample and the post-paradigm 5-minute sample (i.e., 5-minute post – baseline). Like cortisol, there was no difference between groups in sAA reactivity, t(24) = 0.198, p = 0.845, d=0.078. Group$_{Drop}$ (M = -0.804, SE = 1.763) showed similar changes in sAA compared to Group$_{Drop}$ (M = -0.688, SE = 1.152). A recovery difference score was also explored using the baseline and post-paradigm 20-minute sample (i.e., 20-minute post – baseline). Once again, no difference was found between groups, t(23) = -0.640, p = 0.528, d = 0.256 (Group$_{Drop}$: M = -0.412, SE= 1.278, Group$_{NoDrop}$: M = -0.723, SE = 1.155).

E. Discussion.

The primary goal of this project explored individual differences in reactivity and recovery in infants' responses to an established social stressor, the maternal SF, using both behavior and salivary biomarkers. Unlike previous research, infants' stress response was calibrated to the SF by tethering actual infant distress to the timing of post-stress salivary samples rather than relying on an arbitrary time point during the paradigm. This design resulted in two stress categories, infants who exhibited extreme distress and did not complete the paradigm (Group$_{Drop}$) and infants who did not and completed the paradigm (Group$_{NoDrop}$).

Overall, there were no differences in salivary cortisol or sAA between the two stress groups. At first glance, this may imply that tethering infant distress to the timing of the salivary post-paradigm samples was unsuccessful and thereby unnecessary. However, three important conclusions can be made of the data. First, when considering cortisol, only a handful of FFSF studies have found differences in the magnitude of the cortisol response. Instead, studies rely on individual differences by examining group differences commonly referred to as "responders" and "non-responders". However, even then, typically less than 50% of infants are classified as responders (Provenzi, Giusti, & Montirosso, 2016). Whereas, in the current study, approximately 73% of $Group_{Drop}$ infants exhibited a cortisol response. This supports the need for a tethered calibrated social stressor in future research. Second, the presence of the law of initial value (LIV) is rarely addressed in FFSF studies where post-stress measures of cortisol might be restricted by baseline levels as seen with other physiological measures such as heart rate. In the current study, a positive relationship was found between baseline cortisol levels and the peak post-stress levels thereby suggesting that the Law of Initial Values may not need to be considered when analyzing salivary cortisol changes in infants. The findings also highlight the importance of considering baseline values when addressing individual differences in stress reactivity. There was a statistical trend showing that $Group_{Drop}$ infants had greater baseline cortisol and sAA levels and this may have contributed to their longevity in the study. Lastly, additional research using sAA as a measure of arousal following a social stressor is necessary. While the current study did find a difference in sAA levels between samples, the difference did not represent the expected pattern. Typically, when related to baseline levels sAA increases 5-minutes following a stressor and then subsequently decreases (although the recovery time frame is contested). In the current study, like cortisol, there were no group differences. However baseline sAA was greater than both the 5-minute and 20-minute post-stress samples. There was no difference between the two post-paradigm samples. The current findings leave one to wonder whether or not sAA is a valid measure of social stress in young infants. Alternatively, because the reactivity timeline for sAA is shorter (i.e., 5-minutes post-stress) compared to cortisol (i.e., 20-minutes post-stress), perhaps the novel events that occurred before taking the baseline sample (e.g., meeting new people, the new lab environment) had an effect on baseline sAA more than baseline cortisol. This calls into question whether or not the sample should be considered a valid baseline measure and whether change in sAA in the SF is a reliable measure of stress in infants.

In conclusion, the current study was the first known of its kind to calibrate infants' stress response to a social stressor. Although additional research is necessary, the findings question the use of both salivary cortisol and salivary alpha-amylase as measures of social stress in early infancy. More consequentially the results suggest that we have yet to identify appropriate measures of individual differences in infants' stress

reactivity. The finding of a relation between cortisol baseline levels and reactivity contrasts with other behavior and physiologic findings, but suggests that there are indeed individual differences in cortisol functioning. Perhaps another question would be whether the SF paradigm as currently used in the extant literature is a sufficient stressor for uncovering individual differences. A more challenging social stressor might be required, like lengthening the duration of the SF, instead of increasing SF repetitions which may have permitted regulatory recovery by the infant.

References

Adamson, L. B., & Frick, J. E. (2003). The Still Face: A History of a Shared Experimental Paradigm. *Infancy*, *4*(4), 451–473. http://doi.org/10.1207/S15327078IN0404_01

Granger, D. A., Hibel, L. C., Fortunato, C. K., & Kapelewski, C. H. (2009). Medication effects on salivary cortisol: Tactics and strategy to minimize impact in behavioral and developmental science. *Psychoneuroendocrinology*, *34*(10), 1437–1448. http://doi.org/10.1016/j.psyneuen.2009.06.017

Mesman, J., van IJzendoorn, M. H., & Bakermans-Kranenburg, M. J. (2009). The many faces of the Still-Face Paradigm: A review and meta-analysis. *Developmental Review*, *29*(2), 120–162. http://doi.org/10.1016/j.dr.2009.02.001

Provenzi, L., Giusti, L., & Montirosso, R. (2016). Do infants exhibit significant cortisol reactivity to the Face-to-Face Still-Face paradigm? A narrative review and meta-analysis. *Developmental Review*, *42*, 34–55. http://doi.org/10.1016/j.dr.2016.07.001

Tronick, E., Als, H., Adamson, L., Wise, S., & Brazelton, T. B. (1978). The Infant's Response to Entrapment between Contradictory Messages in Face-to-Face Interaction. *Journal of the American Academy of Child Psychiatry*, *17*(1), 1–13. http://doi.org/10.1016/S0002-7138(09)62273-1