Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

_____ _____

Blaire Caroline Pingeton Date

Origins of Infant Temperament in the Fetal Heart?

By

Blaire Caroline Pingeton

Master of Arts

Psychology

_____

Sherryl Goodman, Ph.D.
Advisor

_____

Patricia Brennan, Ph.D.
Committee Member

_____

Yuk Fai Cheong, Ph.D.
Committee Member

Accepted:

_____

Lisa A. Tedesco, Ph.D.
Dean of the James T. Laney School of Graduate Studies

_____

Date

Origins of Infant Temperament in the Fetal Heart?

by

Blaire Pingeton

B.M., New York University, 2013

Advisor: Sherryl Goodman, Ph.D.

An abstract of a thesis submitted to the Faculty of the James T. Laney School of Graduate Studies of Emory University in partial fulfillment of the requirements for the degree of Master of Arts in Psychology, 2019.

Abstract

Origins of Infant Temperament in the Fetal Heart?

By Blaire Pingeton

Infants' temperament qualities are concurrently and prospectively associated with their functioning, into adulthood; thus, research of temperament's precursors is vital for both understanding mechanisms by which temperament buffers and aggravates negative outcomes and developing efficacious preventive interventions. Biological processes involved in reactivity and regulation (i.e. heart rate, heart rate variability) are related to temperament, beginning in infancy. Findings linking fetal heart rate (FHR) and fetal heart rate variability (FHRV) with infant temperament are mixed and limited. The current study built on work assessing the developmental continuity of fetal-infant temperament by correcting methodological shortcomings and testing fetal coupling, a FHR-movement variable indexing parasympathetic nervous system development. Data derived from two longitudinal studies of prenatal/infant development in Atlanta, GA (Emory cohort) and NYC (Columbia cohort). In both studies, fetal data were collected in the 3rd trimester. Temperament data were collected at 3- and 4-months postnatal age, via three factors on the Infant Behavior Questionnaire—Revised (IBQ-R): Surgency (SUR), Negative Affectivity (NA) and Orienting/Regulation (REG). We tested the association between FHR, FHRV, coupling, and infant SUR, NA, and REG. Gestational age (GA) at the time of fetal data collection, GA at birth, and infant sex were tested and controlled for if needed. Results differed by site. In the Emory cohort, FHR was positively associated with REG, but not with NA or SUR; FHRV and coupling were not associated with any temperament variables. In the Columbia cohort, FHR was positively associated with NA, but not SUR or REG; FHRV (negatively) and coupling (positively) were associated with SUR and REG, but not NA. In this prospective, longitudinal study of data from two samples that differed in risks relevant to fetal development and temperament, we found mixed support for the hypotheses: three fetal heart indicators—FHR, FHRV, and coupling—were differentially related to the three infant temperament factors. The pattern of associations between fetal heart measures and infant temperament may suggest that fetal heart indices matter more in the prediction of temperament among pregnant women and infants who, on average, experience less depression and stress and are less economically resourced.

Origins of Infant Temperament in the Fetal Heart?

by

Blaire Pingeton

B.M., New York University, 2013

Advisor: Sherryl Goodman, Ph.D.

A thesis submitted to the Faculty of the James T. Laney School of Graduate Studies of Emory University in partial fulfillment of the requirements for the degree of Master of Arts in Psychology, 2019.

Acknowledgements: We are grateful to Drs. Brennan, Cheong, and Monk for their input and support of this project. We are tremendously grateful for all of the pregnant women and infants who participated in our study! BCP is grateful for the time, guidance, expertise, and mentorship of SHG. Danny Gattas provided vital moral support for this project. The Emory study was funded by P50MH077928 and the Columbia study was funded by 1R01MH092580-01A1; further funding was from NSF 16-588.

Table of Contents

List of Tables

Infants' temperament qualities are both concurrently and prospectively (into adulthood) associated with their behavioral, emotional, and social functioning (Rothbart, 2007; Rothbart, Ahadi, & Evans, 2000; Sanson, Hemphill, & Smart, 2004; Zentner & Bates, 2008). Moreover, certain infants' temperament qualities, such as negative affectivity, are associated with more depression in mothers (Beck, 1996; Britton, 2011; Mayberry & Affonso, 1993; McGrath, Records, & Rice, 2008). Given this evidence for the importance of infants' temperament, both for the infant and the mother, researchers have turned to questions about the origins of temperament. Specifically, might infant temperament originate during fetal development? Knowledge of fetal origins of temperament would support efforts to identify specific polygenetic, neurological, and epigenetic markers associated with temperament factors. Critically, knowledge of fetal origins of temperament, given the adverse outcomes for infants and mothers associated with infant temperament, would support the design and implementation of clinical trials of interventions to buffer these adversities.

Dating back to the 1938 Fels Institute study, researchers have speculated that temperament may have its origins in patterns of fetal heart rate (FHR), fetal heart rate variability (FHRV), and fetal movement (FM). On what conceptual or empirical bases might one expect to see origins of infant temperament in these fetal markers? Support derives from several sources. Two core features of temperament, per Rothbart's definition (Rothbart, 2007), are reactivity and regulation, both of which might be indexed in the fetus by FHR and FHRV. In addition to fetal heart markers, the construct of "coupling" takes into consideration the relationship between FHR and fetal movement. Given the coordination that coupling requires, it is thought to index both parasympathetic nervous system and neurological maturation (DiPietro, Hodgson, Costigan, Hilton, & Johnson, 1996). In these ways, it mirrors postnatal behavior assessment, such as the

Neonatal Behavioral Assessment Scale (NBAS), which concurrently measures temperament constructs (e.g. activity level, irritability, cuddliness, regulation of state) and neurological development (e.g. reflexes, general tone, cost of attention) (Als, Tronick, Lester, & Brazelton, 1977). Yet, although DiPietro et al. (2018) found coupling to be associated with behavioral inhibition in middle childhood, we found no published reports of empirical studies of coupling in relation to infant temperament.

Fetal Heart Rate and Infant Temperament

A main conceptual support for investigating whether fetal heart rate patterns relate to infant temperament derives from the role of the heart in autonomic nervous system (ANS) processes. The ANS may mediate the regulatory and reactivity behaviors that are central to Rothbart's definition of temperament (Derryberry & Rothbart, 1988; Rothbart, 1989). This idea is supported empirically by findings that low resting heart rates are associated with antisocial and violent behaviors (possibly indicating a lack of reactivity or fearless predisposition) in children and adolescents, relative to healthy controls (Ortiz & Raine, 2004). In contrast, a high resting heart rate is associated with inhibition and anxiety in later childhood and adolescence (Kagan, Reznick, & Snidman, 1988; Mezzacappa et al., 1997).

In terms of studies linking FHR and temperament in infants, prospective studies offer some empirical support. Kagan and colleagues (Snidman et al., 1995) examined associations between baseline heart period in fetuses (n = 66) at a mean gestational age of 38.5 weeks with baseline infant heart period measured during sleep at ages 2 weeks, 2, 4, 6, and 14 months and in the lab after exposure to an age appropriate unfamiliar event at 2, 4, 9, 14, and 21 months. For their temperament measure, these researchers categorized infants into one of four reactivity groups (low to high) based on cry and fear scores obtained during observations of the infants

exposed to the unfamiliar events. [Note: heart period is a monotonic, non-linear transformation of heart rate. Some investigators choose to use heart period (rather than heart rate) for statistical or physiological reasons related to their study variables and design (Porges & Byrne, 1992)]. The four reactivity groups did not differ in baseline fetal heart period. However, the low reactivity infants were significantly more likely, relative to the high reactivity infants, to have baseline fetal heart periods above the median value for all fetuses (63% relative to 18%). Among other studies of baseline fetal variables predicting infant temperament, DiPietro et al. (1996), reported that baseline FHR at 36-weeks' gestation (n = 31) was positively associated with unpredictability at 3- and 6-months (r's = .46 and .45 respectively) and unadaptability and dullness at 6-months (but not 3-months) (r's = .34 and .36, respectively). FHR was negatively associated with infant activity level at both 3- and 6-months (pr's = -.39 and -.41, respectively). In sum, these prospective, correlational studies provide support that a baseline fetal heart rate is related to infant temperament characteristics associated with reactivity and motor movement. However, given the two very different approaches to measuring temperament, it is difficult to draw conclusions. Snidman et al. (1995) found higher fetal heart period to be related to low reactivity (low cry and low movement in response to sensory stimuli), whereas in DiPietro et al. (1996), high FHR was associated with more parent rated dullness, unadaptability, and unpredictability, and less activity. Crucially, differences between these studies may be a result of approach to measuring temperament, i.e. laboratory vs. maternal report and the paradigm that Snidman et al. used, which defined reactivity as observed motor movement and crying. Thus an important next step in this line of research would be to examine fetal predictors of infant temperament using standardized approaches to measuring both constructs. Furthermore, the replicability and generalizability of these studies may be limited due to their small sample sizes

and sample demographic homogeneity. With regard to the latter, both Kagan and DiPietro used samples of Caucasian women (100% in Kagan, 81% in DiPietro), most of whom had a college degree.

Further empirical support for an association between FHR and infant temperament comes from a few experimental studies, in which researchers observed FHR in response to either an imposed stressor (mothers completing a Stroop task or watching a labor and delivery video) or took advantage of the naturally occurring stress of labor and delivery. First, DiPietro, Ghera, and Costigan (2008) found that fetuses (n = 137) whose heart rates decreased in response to their mothers' watching an evocative labor and delivery video were more likely to be irritable, operationalized via infant cry response to a series of developmental tasks. Second, Emory, Walker, and Cruz (1982) used six FHR composites (Deceleration, Complexity, Frequency, Baserate, Acceleration, and Variability) to predict proprietary Brazelton factors (Attention-Orientation, Arousal, and Temperament). They found that Baserate (the combination of baserate FHR and average deceleration) was negatively (β = -.33) and statistically related to Attention-Orientation; Deceleration and Acceleration were negatively associated with Attention-Orientation (βs = -.34, -.19) but were not significant in the multiple regression. Third, Werner et al. (2007) found that (n = 50) fetuses whose heart rates increased, but did not decrease, during the Stroop task were move likely to be classified as "high cry reactivity" using Kagan's paradigm at infant age 4 months. Overall, there is no consistent pattern between specific temperament constructs and stress-response FHR. Further, Werner et al. (2007) and DiPietro et al. (2008) find FHR response predictive of reactivity, but in opposite directions; this may be due to differences in the prenatal stressors employed, and/or to the use of a nonstandard temperament instrument (in the case of DiPietro et al, 2007).

In summary, there are empirical findings linking FHR, at both baseline and after a stressor, to infant temperament. However, across studies there is little to no consistency among which temperament constructs are related to FHR, the effect sizes of the associations, or the direction of associations. There are additional concerns related to the typically small sample sizes and the unstandardized nature of some of the measurement instruments of both FHR and infant temperament. Further, with one exception, in all of the studies of FHR and infant temperament, the study sample demographic characteristics are quite homogenous and participating pregnant women were primarily Caucasian, adult, well-educated, and financially secure. These sample characteristics are concerning since several studies have found that there are differences between low- and middle-income families in both fetal development and infant temperament (Conger, Conger, & Martin, 2010; Lantz, House, Mero, & Williams, 2005; Rothwell & Han, 2010). Thus, reliance on homogenous, highly resourced study samples may limit generalizability regarding the association between fetal characteristics and infant temperament, leaving unknown the extent to which fetal characteristics and infant temperament are associated in low-income, more ethnically diverse samples.

Fetal Heart Rate Variability and Infant Temperament

Researchers conceptualize FHRV as indexing regulation. Broadly speaking, HRV measures two competing systems in the autonomic nervous system: the excitatory sympathetic nervous system (SNS) and the inhibitory parasympathetic nervous system (PNS). Stress elicits an SNS response characterized by increased heart rate; during periods of relaxation the PNS lowers the heart rate and decreases physiological arousal. Heart rate variability is a measure of an individual's ability to regulate physiological reactivity in response to stress by modulating between these two systems.

Consistent with this conceptualization, a few published studies offer empirical support for an association between FHRV, measured both at baseline and reactivity, and infant temperament. In a small sample (n = 31), DiPietro et al. (1996) found that baseline FHRV at 36-weeks gestational age was positively associated with infant activity level at 3-months (r = .31), and not significantly, but negatively associated with 3- and 6-month fussiness (r = -.05 and r = -.20), 6-month unadaptability (r = -.18) and 3- and 6- month dullness (r = -.25 and r = -.22). In addition to a concern with the small sample size, it is worth noting that there was a relatively small variance in values of FHRV M = 5.77, SD = 1.31.

DiPietro et al. (2008) also contributed an experimental study of FHRV (reactivity), wherein they exposed a group of pregnant women at 32 weeks gestational age (n = 137) to an emotionally charged video about childbirth and measured their fetuses' heart response to this film and calculated FHRV suppression in response to the video. Then, from the FHRV score, they predicted infant temperament reactivity at 6-weeks, albeit with an unstandardized temperament measure, with which they dichotomized infants as either "irritable" or "not irritable." FHRV suppression in response to the video significantly predicted infant irritability classification, $F_{(1,98)} = 4.02$, $p = .01$, $\eta^2 = .039$.

The FHRV and FHR results come from the same studies; hence, previously stated concerns about sample homogeneity persist. In this small literature, one study reports that baseline FHRV is positively associated with infant activity and negatively associated with infant fussiness, unadaptability, and dullness. A separate study demonstrates that infants who show FHRV suppression after a maternal stressor are more likely to be classified as irritable, operationalized as infant reactivity to a developmental task. Replication of these findings—that baseline FHRV is negatively associated with infant emotional reactivity and that a decrease in

FHRV post-stressor increases the likelihood of infant irritability—are needed with modern, standardized, reliable infant temperament instruments.

Fetal Coupling and Infant Temperament

Finally, we found no published studies reporting on the association between fetal coupling and infant temperament, despite clear conceptual links between the two constructs. Fetal coupling is an index of the relationship between fetal heart rate change in response to fetal movement; coupling can be measured in terms of frequency (the number of coupling instances) or latency (the amount of time between a change in heart rate and movement). Across gestation, the latency between movement and a heart rate change shortens, and the number of instances of coupling increases, signaling central nervous system maturation (DiPietro, Hodgson, Costigan, Hilton, et al., 1996). Furthermore, coupling has been associated with development of the parasympathetic nervous system (PNS). The PNS plays a central role in the development of temperament; specifically, higher levels of parasympathetic nervous system activity are related to better emotional regulation in infants, toddlers, children and adults (Beauchaine, 2001; Kreibig, 2010; Mezzacappa et al., 1997; Thompson, Lewis, & Calkins, 2008).

On these bases, we propose that an examination of the relationship between fetal coupling and infant temperament would serve several purposes. First, investigating the association between fetal coupling and infant temperament would provide information about how early (i.e. during pregnancy) the association between parasympathetic control (indexed by fetal coupling) and emotion regulation emerges. Second, coupling would provide an additional index of fetal parasympathetic control and its relationship to infant temperament over and above FHRV. Third, given the relationship between fetal coupling and CNS development, an examination of the coupling-temperament relationship could clue investigators as to whether or

not it is general fetal neurological development that is driving the association between fetal markers and infant temperament. Though none of these claims can be answered conclusively using fetal coupling, a strong association between coupling and fetal variables would warrant further investigation, which could substantially augment our understanding of the mechanisms underlying the etiology of temperament.

The Current Study

The current study aimed to replicate and expand upon previous work assessing the developmental continuity of fetal-infant temperament. First, we attempted to correct for methodological shortcomings. In particular, when available, we relied on large samples to test our hypotheses and used well-established measures of fetal functioning and temperament. Second, we tested the association in two samples both of which were at risk of atypical fetal functioning and infant temperament, to assess whether developmental continuity between the fetal heart and infant temperament replicated across both samples. One sample was demographically typical of previous studies, i.e. predominantly white, well-educated, middle income or higher, but at elevated risk of depression during pregnancy and postnatally, because of a history of past depressive episodes. The other sample was an ethnically and socioeconomically diverse, community sample. Employing these two diverse and distinct cohorts allowed us to test the generalizability of the hypothesized fetal origins of temperament. Further, if the two samples are found to not differ on key variables, we planned to merge the samples and, thus, enhance our sample size. Third, we tested the relationship between a fetal heart rate-movement coupling variable and infant temperament. These three aims dovetailed with our broader goal of furthering the understanding of fetal origins of temperament in a variety of environmental contexts in order to inform the design of interventions to enhance children's social and emotional development.

We relied on Rothbart's Infant Behavior Questionnaire-Revised (IBQ-R) for measurement of infant temperament (Gartstein & Rothbart, 2003). We selected the IBQ-R based on its demonstrated cross-cultural reliability and validity and its reliance on parent report. Further, the IBQ-R is widely held to be one of the most studied and best-constructed parent report measures of infant temperament (Stifter & Dollar, 2016). There are strong arguments on both sides for the use of parent report versus laboratory assessment of infant temperament. We chose to use a parent report instrument because parents can report on an infant's typical behavior in a variety of contexts; this contrasts most studies of the fetal origins of infant temperament, which relied on a single laboratory visit for postnatal temperament assessment. Items on the IBQ-R load onto 14 subscales, which in turn load onto three factors: Surgency/Extraversion, Negative Affectivity, and Orienting/Regulation. Using these broad factors contrasted previous work, which measured more narrowly defined temperament constructs (e.g. irritability, cry reactivity).

With respect to the timing of our assessments, we aimed to assess infant temperament as early and reliably as possible based on the literature and our study instrument's capabilities. We measured the fetuses as late as possible in the 3rd trimester (GA 28 weeks – birth) based on the literature, which found associations between 3rd trimester fetal variables and infant temperament (DiPietro, Ghera, & Costigan, 2008; Werner et al., 2007). We assessed infant temperament at 3- or 4-months postnatal age, to capitalize on the earliest age that the IBQ-R can validly measure (Gartstein & Rothbart, 2003). This gap best allowed us maximize confidence in our temperament measurement while constraining the influence of postnatal developmental factors.

Given mixed findings from studies of FHR and FHRV at either baseline or in response to a stressor, we had no empirical support for a hypotheses about whether baseline fetal heart

measures or measures of the fetal heart post-stressor would better predict infant temperament. Thus we relied on an understanding of temperament to derive this hypothesis. That is, given that temperament captures global, trait-level functioning, rather than behavior in an individual circumstance (Rothbart, Derryberry, & Hershey, 2000; Shiner & Caspi, 2003), we expected that baseline fetal heart measures, capturing a fetus' "typical" behavior, would be predictive of temperament. On the other hand, given that reactivity is a core feature of temperament, post-stressor reactivity could be a more ecologically valid index of an infant's eventual temperament. On this basis, we also expected post-stressor fetal measures to show a relationship to infant temperament.

Related to our aim of gaining greater clarity on the associations between fetal functioning and infant temperament, we included measures of prenatal perceived stress and depression symptom levels (S. Cohen, Kamarck, & Mermelstein, 1994; Cox, Holden, & Sagovsky, 1987). We chose to report psychological distress variables because stress and depression are known to influence FHR, FHRV, coupling, and infant temperament (Austin, Hadzi-Pavlovic, Leader, Saint, & Parker, 2005; Buitelaar, Huizink, Mulder, de Medina, & Visser, 2003; Monk et al., 2011; Monk et al., 2004). Despite these associations, published studies of fetal predictors of infant temperament have not reported findings on participants' depression or stress. Thus, we characterized maternal stress and depression during pregnancy to test whether associations between fetal markers and infant temperament differ based on mothers' levels of depressive symptomatology and perceived stress.

Our specific hypotheses were based on theory as much as on the empirical literature; that is, how we hypothesized FHR, FHRV and coupling theoretically related to the infant temperament constructs of Surgency, Negative Affectivity, and Orienting/Regulation. We chose

this approach because of the mixed findings from the literature testing the association between FHR and FHRV and infant temperament, and the absence of published studies of fetal coupling and infant temperament. Additionally, aforementioned concerns regarding measurement and sampling reduced our confidence in the findings from the published studies generalizing to at-risk populations such as the two we sampled.

Due to the strong empirical relationship between heart functioning and emotional expression, we constrained our FHR and FHRV hypotheses to infant temperament measures of emotionality. Specifically, we hypothesized that FHR would be positively associated with Surgency and Negative Affectivity, since a higher HR is associated with ANS reactivity. Similarly, we hypothesized that FHRV would be negatively associated with Surgency and Negative Affectivity, given knowledge that more parasympathetic activation is associated with better emotional regulation (Beauchaine, 2001; Kop et al., 2011; Stifter, Dollar, & Cipriano, 2011). The studies testing associations between FHR and FHRV and the constructs measured by Orientation/Regulation yielded mixed findings; therefore, we did not specify a hypothesized direction of associations between these variables but tested them in an exploratory manner. In terms of coupling frequency, as an index of CNS development, we hypothesized that it would be positively associated with Regulation/Orienting. We did not specify a direction of the association between coupling and Surgency or Negative Affectivity, since we could not conclusively predict how coupling's relationship to both parasympathetic and CNS development would influence coupling's function within these constructs. Overall, we predicted that our hypothesized associations would maintain for both baseline and post-stressor data. Though we constrained our a priori hypotheses to specific, theoretical relationships between fetal characteristics and infant temperament variables, echoing other researchers (DiPietro, Hodgson, Costigan, & Johnson,

1996; DiPietro, Voegtline, Pater, & Costigan, 2018) we planned on testing and reporting each fetal variable with each infant temperament characteristic. We believe that this approach maximizes our contribution to the understanding of the fetal origins of infant temperament due to an extant lack of empirical consensus, the strengths of our study measures, and our unique sample characteristics.

Method

Participants and Study Overviews

Data were derived from two longitudinal studies of prenatal/infant development. One study, hereafter referred to as the Emory study, was conducted at the Women's Mental Health Program at the Emory University School of Medicine and the Department of Psychology at Emory University, in Atlanta, GA. Pregnant women with lifetime histories of mental illness participating in a longitudinal investigation of the impact of perinatal maternal mood and stress on child neurobehavioral outcomes were screened for inclusion. The measures and procedures for the current study were a subset of a larger set of measures and procedures. The fetal monitoring session for the current analyses occurred in the third trimester (gestation age > 28 weeks). At infant age 3-months, the mothers completed the temperament measure. All data collected were coded with a HIPAA compliant identifier and entered into a centralized database. The Emory University Institutional Review Board approved the study, and all women provided informed consent.

The final two samples consisted of n = 160 women (out of n = 275 total) who had baseline 3rd trimester fetal session data and 3-month infant temperament data and n = 157 women who had post-stressor 3rd trimester fetal session data and 3-month temperament. There were no significant differences between women included in this sample relative to women

excluded in terms of age, education level, Hollingshead, gravidity, parity, infant gender, or maximum prenatal depression or stress score. The gestational age of participants not included in this sample was approximately one week and two days shorter (M = 38.52(1.72)) than for participants included in the study M = 37.23(2.46), M difference = -1.29, d = .61, $t_{(472)}$ = -4.67, p < .001. Additionally, more women in the included sample were Caucasian (91% vs. 55%) V = .45, $\chi^2_{(5)}$ = 73.60, p < .001.

The other study, hereafter referred to as the Columbia study, was conducted in the Department of Psychiatry, Division of Behavioral Medicine at Columbia University Medical Center. Women were recruited through the Department of Obstetrics and Gynecology at Columbia University Medical Center for a study assessing the effects of prenatal stress on epigenetic markers in placental tissue. Women were enrolled between 8 – 26 weeks and made two lab visits per trimester. Participants completed two fetal assessments at approximately 24-27 weeks GA and 34-37 weeks GA and were visited in the hospital 12-36 hours after birth by study personnel. Participants were invited to participate in a follow-up study at 4-months postnatal age, which consisted of 1-2 lab visits. The New York State Psychiatric Institute Institutional Review Board approved the original and follow-up study and all women provided informed consent.

The final two samples consisted of n = 78 women (out of n = 187 total) who had baseline 3rd trimester fetal data and 4-month infant temperament data and n = 41 women who had reactivity 3rd trimester fetal data and 4-month infant temperament measurements. There were no statistically significant differences between the included and excluded participants in terms of gravidity, parity, Medicaid status, marriage status, infant sex, or maximum prenatal depression or stress score. Women were approximately 2 years older in the group included in analyses, 30.80

(6.61) vs. 28.81(5.86), $d = .31$, $t_{(185)} = -2.17$, $p = .03$) and fewer women in the included sample were Hispanic/Latina relative to the larger sample (59% vs. 76%, $V = .18$, $\chi^2 = 6.27$, $p = .01$).

Both studies included healthy, pregnant women, 18-45, who did not smoke, drink, or use recreational drugs during their pregnancy and who could read and write English fluently. For the purposes of these analyses, women carrying more than one fetus (twins, triplets, etc.) were excluded. In the Columbia cohort, women had to plan to deliver at a hospital associated with Columbia University Medical Center and could not be taking psychotropic medications at the time of enrollment. In addition, women were excluded if they had a diagnosis of Bipolar Disorder. In the Emory sample, selective serotonin reuptake inhibitor (SSRI) use was not an exclusionary criterion. Crucial to this decision, previous work with this sample has demonstrated that prenatal SSRI status did not influence FHR, FHRV, or coupling values (Gustafsson et al., 2018). Work by others (Gentile & Galbally, 2011; Reebye, Morison, Panikkar, Misri, & Grunau, 2002) has failed to show any relationship between prenatal SSRI use and maternally reported infant temperament.

Fetal Data Collection

Fetal data collection for both sites was designed and overseen by Dr. Monk, as described in previous work (Doyle et al., 2015; Gustafsson et al., 2018; Werner et al., 2007). Procedures and equipment were identical at the two sites with two exceptions. One, at the Emory site a single, 5-minute baseline was administered, whereas at the Columbia site two baselines (first a 20-minute and then a 5-minute) were administered. Second, while both sites administered the Stroop color-word matching task as a stressor, Emory's second stressor task was an arithmetic challenge, while Columbia's second stressor task was a breathing challenge. For hypothesis testing, we relied on the 20-minute baseline, given that it was administered first and that the fetal

variables are calculated as a mean, thus enhancing the likelihood that the two procedures could be considered as equivalent. At both sites, baseline was followed by two stressor tasks (5-minutes each), with a 5-minute recovery period after each stressor. Data were obtained using a Toitu MT 325 fetal actocardiograph (Toitu Co., Ltd, Tokyo, Japan), which detects FM and FHR via a single transabdominal Doppler transducer. The fetal data were collected from the Toitu's output port, digitized at 50 Hz using a 16–bit A/D card (National Instruments 16XE50) and analyzed offline. Separately, for both baseline and post-stressor (the recovery period after a stressor), FHR was calculated as the mean value for the study period. FHRV was computed by calculating the standard deviation of the FHR. Coupling, operationalized here as the frequency of instances of fetal heart rate and fetal movement synchrony (described in detail in Doyle et al., 2015) was computed in overlapping 4-minute segments; artifacts were removed and the average of these segments was computed for each paradigm period. Fetal sessions from the 3rd trimester were used for analyses; when there was more than one fetal session collected in the 3rd trimester, the latest session was use for analyses.

Infant Temperament

In both sites, infant temperament was derived using the 191-item parental report instrument, Infant Behavior Questionnaire—Revised (IBQ-R) (Gartstein & Rothbart, 2003), completed by the mothers. The IBQ-R instructs parents to rate their infant's behavior during the past week in a variety of domains on a seven-point scale, from one (Never) to seven (Always). The questionnaire yields scores on 14 scales, with ten to 18 items per scale and scale scores being the mean of items on that scale, such that higher scores indicate more of the measured temperament characteristic. Scales cluster into three overarching factor scores: Surgency/Extraversion (SUR), Negative Affectivity (NA), and Orienting/Regulatory Capacity

(REG). Surgency is a measure of positive emotionality; the subscales that comprise the factor are High Intensity Pleasure, Activity Level, Impulsivity, and reversed Shyness. Negative Affectivity is a measure of negative emotionality; the subscales that comprise the factor are Sadness, Discomfort, Anger/Frustration, Fear, and reversed Falling Reactivity/Soothability. Orienting/Regulation is a measure of effortful control; the subscales that comprise the factor are Low Intensity Pleasure, Inhibitory Control, Attentional Focusing, and Perceptual Sensitivity. The IBQ-R's reliability and validity have been demonstrated through monomethod discriminant validity, a demonstrated similar factor structure between the IBQ-R and later methods of temperament assessment (such as the Child Behavior Checklist and adult temperament questionnaires), convergence with laboratory observation, and modest inter-rater reliability between caregivers (Gartstein & Rothbart, 2003; Parade & Leerkes, 2008).

Maternal Characteristics and Birth Outcomes

In both sites, women self-reported their demographic characteristics during the first study session. Women gave permission for access to their electronic medical records on labor and delivery, from which infant birthweight and gestational age (GA) at delivery were abstracted. Since we did not constrict the fetal assessment to a window within the 3rd trimester as others have done, the timing of the fetal session (in weeks gestation) was included as a covariate, to rule out variability that could be a result of gestational age at the time of the fetal session. GA at the time of birth and infant sex were also tested as covariates.

Psychological Variables

To characterize maternal psychological functioning during pregnancy, we measured depression symptom severity and perceived stress. Perceived stress was measured using the 14-item self-report instrument the Perceived Stress Scale (PSS) (S. Cohen et al., 1994). The

instrument measures the extent to which respondents perceived events during the last one-month period to be “unpredictable, uncontrollable, and overloading”. The PSS is a reliable measure (coefficient alpha when the test-retest interval is two days = .85), and demonstrates concurrent and predictive validity through its relationship to life events, physical symptoms, depressive symptoms, and health. The PSS has been validated in several multi-cultural pregnant samples (Ramírez & Hernández, 2007; Remor, 2006; Siqueira Reis, Ferreira Hino, & Romélio Rodriguez Añez, 2010). Prenatal depression symptom severity was measured using the Edinburgh Perinatal Depression Scale (EPDS), a widely used, 10-item, self-report measure (Cox et al., 1987). The questionnaire has high internal reliability and good validity (Cox et al., 1987). For both variables, we relied on the peak or maximum prenatal score for the analyses.

Data Analytic Plan

All analyses were performed in SPSS (Corporation, 2016). Descriptive statistics were calculated using the mean and standard deviation for continuous variables and percentage for categorical variables. Analyses were conducted pooled and separately, when appropriate. For pooled analyses a dummy variable was created to test and measure site effects and infant sex. A stepwise multiple regression was run for each of the three DV infant temperament factors (Surgency, Negative Affectivity, and Orienting/Regulation). Step one included the covariates and site variable (GA at the time of the fetal session, GA at the time of birth, and site); step two included the fetal measures (FHR, FHRV, and coupling) at either baseline or post-stressor. After the full model was run, covariates not significantly associated with the temperament factor were removed and the model was re-run. Then, fetal measures not associated with the temperament factor ($p > .1$) were removed and the model was run again. Procedures for analyses separately by site were identical, except a dummy variable for site effects was not included at any stage. All

independent variables were centered. Effect sizes are reported using Cohen conventions (J. Cohen, 1992).

Results

Sample Characteristics

See Table 1 for a description of study characteristics divided by site, with effect sizes for site comparisons when data were available.

Emory

See Table 1 for sample descriptive statistics. The Emory sample consisted of $n = 160$ pregnant women with baseline fetal data and temperament data; 157 (98%) of those women had post-stressor fetal data. Most women identified as Caucasian. The mean Hollingshead score indicates that on average women in this sample had incomes in the middle to upper-middle range. There were no statistically significant differences between the total sample and women with both baseline and post-stressor data on any demographic variables (p's > .95).

Columbia

See Table 1 for sample descriptive statistics. The Columbia sample consisted of $n = 78$ pregnant women with baseline fetal data and temperament data, 41 (53%) of whom had post-stressor data. Most women identified as Hispanic/Latina. About half (43%) of the women were receiving Medicaid. There were no statistically significant differences between the total sample and women with both baseline and post-stressor data on any demographic variables (p's > .5).

Site comparison

As shown in Table 1, comparing between sites, Emory women had more economic resources. Emory women were on average about 3 years older than the Columbia women, had on average 1.2 years more education, and were more likely to be married. [Note: there was one

outlier in the Columbia sample, with 28 years education. When this participant was removed, the mean level of education in the Columbia dropped to 15.22 years, and the mean difference increased to 1.36. Since the study was conducted within an academic medical center, we do not suspect measurement error for this value].

There were no difference in gravidity or parity between sites. Babies in the Columbia cohort were born approximately one week later (39.35 weeks) then babies in the Emory cohort (38.52 weeks).

Psychological Characteristics

See Table 1 for maximum prenatal depression and stress scores values, and for a comparison of scores between sites. To account for the considerable heterogeneity of cut-off scores used in the literature, two values indicative of "probable depression" were reported for interpreting the EPDS (Matthey, Henshaw, Elliott, & Barnett, 2006). A score of ≥10 takes a highly sensitive approach; a score of ≥15 takes a highly specific approach. The mean maximum prenatal PSS and EPDS scores were higher for Emory women. In the Emory sample, using a sensitive cut-score, most women were probably depressed (55%), using a specific cut-score, approximately one-third of women were depressed (29%) during pregnancy. In the Columbia sample, using the 10 or more cut-off score (highly sensitive) 22% of participants scored as "probably depressed"; using the 15 or more cut-off score on the EPDS (highly specific), 5% of participants scored as "probably depressed". However, the EPDS was added as a study measure halfway through data collection, thus, data exists for only $n = 46$ (59%) of the sample.

Fetal Measures

Baseline vs. Post-Stressor

A t-test comparing baseline and post-stressor fetal measures revealed no significant differences. As previously reported, 157 (98%) women had post-stressor data in the Emory sample, while only 41 (53%) women in the Columbia sample had post-stressor data. We constrained our analyses to the baseline period for four reasons. First, we had no a priori hypotheses distinguishing between baseline and post-stressor results. Second, the baseline period contained more study participants. Third, an aim of the study was to conduct well-powered tests of the predicated associations, to potentially correct for findings in the literature from studies with small sample sizes. Fourth, we wished to reduce possible type II error. Results for fetal-infant regressions using post-stressor data and t-test results are presented in Appendix A. Otherwise all results presented in this manuscript refer to fetal measures collected during the baseline study period.

Site comparison

See Table 3 for the sample size, range, mean, and SD for the fetal measures at each site and pooled across sites. There were no differences between sites for the FHR and FHRV variable. There was a small difference between sites for one of the three fetal measures. Columbia babies showed slightly more coupling, $d = .29$, $SE = .01$, $p < .05$.

Sex Differences

There were no sex differences for any of the three fetal variables (Table 4).

Infant Temperament Measures

The means and standard deviations of all three temperament factors are presented in Table 5, separately by site and pooled across sites. There were significant site differences for both the Negative Affectivity ($d = .39$, $p = .008$) and Orienting/Regulation factors $d = 1.18$, $p <$

.001. The Emory infants were higher on Negative Affectivity and Orienting/Regulation relative to the Columbia infants.

Sex Differences

There were no sex differences for any of the three temperament factor scores (Table 4).

Fetal-Infant Associations

The pattern of correlations across the fetal, temperament, and potential control variables are shown in Tables 6. The full and final models are presented for each temperament factor in Tables 7-18. Since infant sex was not associated with any fetal or temperament measure, it was not included in analyses.

Surgency

Gestational age (GA) at birth, GA at the time of the fetal session and site were not associated with the Surgency temperament factor ($p > .1$) In the Emory cohort, there were no associations between fetal variables and SUR. In the Columbia cohort, coupling was positively associated with SUR and FHRV was negatively associated at the trend level; FHR was not associated with SUR. When the cohorts were pooled together, coupling was positively associated with Surgency; FHR and FHRV were not associated with SUR.

Negative Affectivity

To account for a site effect, analyses were only conducted separately. GA at birth and GA at the fetal session were not associated with NA. In the Emory cohort, FHR, FHRV, and coupling were not associated with NA. In the Columbia cohort, FHR (positively) accounted for a moderate proportion of the variance in NA scores; FHRV and coupling were not associated with NA.

Orienting/Regulation

To account for a strong site effect, analyses were only conducted separately. In the Emory cohort, FHR was positively associated with REG. FHRV and coupling were not associated with REG. In the Columbia cohort, the three fetal markers accounted for 14% of the variance in infant Regulation/Orienting, a strong effect. Independently, coupling was positively associated with REG; the association between FHRV and REG was negative and not significant (p = .05). FHR and REG were not associated.

Discussion

In this prospective, longitudinal study, three baseline fetal heart indicators—fetal heart rate, fetal heart rate variability, and coupling—were differentially related to the infant temperament constructs of Surgency/Extraversion, Negative Affectivity, and Orienting/Regulation. More associations between fetal heart measures and infant temperament were found in the Columbia cohort. For these women and infants, FHRV was negatively, as hypothesized, associated with Surgency; in exploratory analyses coupling was positively associated with Surgency. We hypothesized that FHR was be positively associated with Surgency, but found no relationship between FHR and SUR. As hypothesized, FHR was positively associated with Negative Affectivity; contrary to hypotheses, FHRV was not associated with Negative Affectivity. Our exploratory analyses found no relationship between coupling and Negative Affectivity. As hypothesized, coupling was positively associated with Orienting/Regulation; exploratory analyses showed a negative relationship between FHRV and REG and no relationship between FHR and REG. In the Emory cohort, exploratory analyses showed that FHR (positively) was associated with Orienting/Regulation; no other associations between fetal heart measures and infant temperament were found.

When observed, the associations between fetal markers and infant temperament constructs were in the hypothesized direction, with the exception of FHRV. FHRV was negatively associated with Surgency and Regulation/Orienting in the Columbia sample at the trend level (there were no associations between FHRV and any infant temperament marker in the Emory sample). We hypothesized that FHRV would be negatively associated with Surgency and NA, but did not anticipate a relationship between FHRV and Regulation/Orienting. One possible explanation of this paradoxical effect may be that one of the scales that makes up the Regulation/Orienting factor is "Low Intensity Pleasure" which measures an infant's enjoyment during low-intensity activities or in relatively unstimulating environments (i.e. "When playing quietly with one of his/her favorite toys, how often did the baby show pleasure") (Gartstein & Rothbart, 2003). Within this sample, it may be that infants low on FHRV are less regulated even in calm environments. To test this, targeted post hoc analyses of the scales (rather than factors) should be conducted.

We predicted that fetal coupling would be related to Regulation/Orienting in both samples, on the theoretical basis of parasympathetic maturation. That is, since coupling requires coordination between the motor and cardiovascular systems, fetuses capable of exhibiting coupling behavior more frequently in utero may, as infants, be better able to coordinate the multiple systems involved in attention and regulation. This interpretation is consistent with DiPietro et al.'s (2010) finding that fetuses who have more coupling scored better on a brainstem auditory evoked potential (BAEP), although not on the Dubowitz Neurological Exam. The positive relationship between coupling and BAEP suggests a second interpretation, that increased coupling is a reflection of increased central nervous system development. Coupling is thought to be controlled by the central nervous system sending parallel signals to the cardiovascular and

motor systems (DiPietro, Costigan, & Voegtline, 2015); postnatally, this increased general CNS development may be responsible for infants' higher scores on REG measures. However, in the same DiPietro et al. (2010) study, greater FHRV was also associated with more optimal BAEP outcomes, whereas in the Columbia cohort of our study, FHRV was negatively associated with REG. Therefore, coupling's, and not FHRV's association with Orienting/Regulation may reflect a specific factor of parasympathetic control rather than general neurological development. To more conclusively determine the function of parasympathetic and CNS development in these associations, future work should test neurological and temperament indices together within the same sample.

Our finding of FHR as positively associated with NA in one of our samples (Columbia) somewhat aligns with work by DiPietro et al. (1996) and Snidman et al. (1995). That is, we replicated DiPietro et al.'s (1996) finding that FHR was positively associated with infant unpredictability (most similar to IBQ-R Negative Affectivity), but could not replicate her finding on FHR being related to activity level (a subscale of the IBQ-R's Surgency factor). It is difficult to compare Snidman et al.'s (1995) findings to ours given the operationalization of temperament in each case; Snidman et al. (1995) found that higher (above the sample median) FHR was related to reduced 4-month reactivity. In the Columbia cohort, FHR was positively associated with NA. Most of the subscales of the NA factor are positively related to reactivity constructs (e.g. fear, discomfort), which contrasts Snidman et al.'s (1995) finding. However, one subscale that loads negatively on the NA factor is Falling Reactivity/Soothability. This negative association with NA mirrors Snidman et al.'s (1995) study. Additionally, our finding in the Emory cohort that FHR was positively associated with REG somewhat mirrors Snidman et al.'s findings. The subscales of the REG factor are related to constructs which measure an infant's

regulatory functioning (e.g. Duration of Orienting, Soothability), which may make these babies less reactive, echoing Snidman et al. (1995). Of note, the Emory sample was more demographically similar to both of these studies. From the same DiPietro et al. (1996) study, the most robust finding with respect to FHRV was in relation to 3-month infant activity level, which the factor structure of our infant temperament assessment measure does not allow us to test. However, non-significant associations in DiPietro et al.'s (1996) study showing a negative relationship between FHRV and dullness somewhat mirror our findings regarding FHRV and SUR. In sum, it is difficult to compare our studies to the extant literature, since previous worked defined temperament with more narrowly than our study, which utilized factor-level variables. A next step to make this comparison more meaningful may be to analyze particular IBQ-R subscales of interest.

While there were minimal differences between sites in fetal heart measures, there were important differences in sample demographics, psychological characteristics, and infant temperament. The women in the Emory sample, all of whom had a history of major depressive episodes, were predominantly affluent, well-educated, and Caucasian. Despite their sociodemographic advantages, relative to the Columbia cohort, they had higher levels of depression and stress prenatally, consistent with their depression risk status. In contrast, almost half of the women's earnings in the Columbia sample qualified them for Medicaid, and most women in the Columbia sample had not obtained a Bachelor's degree. Furthermore, most women in the Columbia sample were Latina, a risk factor for prenatal stress and postpartum depression (Liu, Giallo, Doan, Seidman, & Tronick, 2016). Despite this constellation of sociodemographic risk factors, the Columbia sample experienced lower levels of prenatal stress and depression than the Emory sample. On average, the Emory infants, relative to the Columbia infant, were higher

on Negative Affectivity and Regulation/Orienting, although not significantly different on Surgency. These sample characteristics provide an important frame for the pattern of results observed between samples. The most salient difference between samples is in volume: the Columbia sample had many more fetal-infant associations than in the Emory sample.

Putting together the differences in fetal-infant continuity between the Columbia and Emory cohorts and the fetal-infant continuity found in the literature, one interpretation of our findings is that postnatal factors influence infant temperamental development. With the exception of Orienting/Regulation, the differences between groups on both fetal and infant temperament measures are few and small; however, the associations between these measures do vary between cohorts. It is possible that the depression and stress that the Emory fetuses are exposed to in utero renders them more differentially susceptible to postnatal influences. In this interpretation, one could conceptualize that a relationship between fetal heart measures and infant temperament is normative, and therefore intact in the Columbia sample. However, this normative developmental pathway is disrupted by high levels of prenatal stress and depression, and therefore does not exist in the Emory sample. Further, when situating the results of this study within the broader literature, it is clear that the particular fetal measures associated with infant temperament constructs vary dramatically between samples, even when fetal collection materials and infant temperament measures are held constant across groups. This may mean that biological predisposition for temperament is context-dependent. In this interpretation, specific teratogens or environments (such as depression in our current study) may influence how deterministic an infant's prenatal biological characteristics are in their postnatal environment. An important future direction in this project, spawn by this theory, will be testing whether

prenatal depression or stress, or postnatal parenting sensitivity moderates the association between fetal variables and infant temperament.

Our study contributes to the understanding of the putative fetal origins of infant temperament in many unique ways. One important contribution of this paper is that it tests the continuity of fetal heart measures to infant temperament in two at-risk samples, previously unrepresented in this literature. As already noted, differences between study sites are possibly indicative of environmental agents (in this case, maternal psychological characteristics) which promote differential plasticity. Our study may also assist investigators in designing fetal data collection protocols, since we found that the timing of the fetal procedure (in gestational weeks) did not significantly contribute to the association between fetal heart indices and infant temperament. This is an important contribution to this literature, given the heterogeneity of gestational ages that researchers have tested. In our sample, gestational timing of the fetal session was not related to any combinations of fetal marker and infant temperament factor. Finally, the use of broad factors in measuring infant temperament is a novel contribution of this study. Previous work had focused on narrow constructs within infant temperament. Our study may be more comparable to future work tracking this association throughout postnatal development, since at least two of the three factors (NA and SUR) used in our study are measurable in later childhood and adolescence (Gartstein & Rothbart, 2003).

There are a number of limitations to our study that should be noted when interpreting our results. To begin, though our total sample was much larger than any other studies that have previously tested this association, the sample size of the Columbia cohort was modest, though still larger than most other samples in the literature (n = 78). Furthermore, one of our descriptive statistics (maximum prenatal depression) had considerable “missing completely at random” data,

due to a study design decision. Since the data were truly missing completely at random, and since this variable was used purely for descriptive statistics, we do not believe that it influences our results or interpretations. Another limitation of our study relates to differences in the fetal protocol at each site. The Columbia cohort, relative to the Emory cohort, had an extra twenty minutes of baseline data collected during the session protocol; thus, the fetuses had more opportunity to move out of range of the data collection instruments, which might explain the post-stressor data being available on only about half of those with baseline data in the Columbia cohort. Finally, as DiPietro et al. (2018) note, while post-stressor measures have an alluring conceptual relationship to temperament constructs as defined by Rothbart and others, relying on a maternal stressor may be problematic when one considers the heterogeneity of uterine environments and maternal physiological stress response phenotypes. The wide variety of combinations of womb thickness and fluid type by maternal heart rate and respiratory response likely makes the experimental stressor inequivalent across dyads. Future work should design study protocols that preserve fetal signals through this study period, establishes equivalent stressors across study participants, and more specifically captures reactivity. There is a limitation related to the timing of the administration of our dependent variable. The Emory sample completed their rating scales approximately one month before the Columbia sample. Normatively, there is a linear decrease in Negative Affectivity and an increase in Regulation/Orienting. Thus, while the differences in Negative Affectivity may be a result of postnatal age, the differences in Regulation/Orienting may be underestimated due to the month age gap in measurement between sites. In the Emory cohort women filled out this measure at 3-months postpartum age; in the Columbia cohort women completed the IBQ-R at 4-months postpartum age. While the IBQ-R creators do not note any differences between 3- and 4-months

in any temperament dimensions, they do caution that certain constructs and scales (i.e. Fear) change from early to late infancy. A future direction will be conducting sensitivity analyses to compare infants from both groups whose ages overlap. Finally, the current study does not contain a "control group". Put a different way, both of our samples were at-risk for different reasons, and we did not compare our samples against a sample of well-resourced women without a history of mental illness. However, the literature has already documented the association between fetal markers and infant temperament in middle-income, Caucasian, mature dyads.

Our study indicates that in a diverse sample using well-validated measures there are prenatal antecedents to infant temperament. However, contextual factors may influence which biological markers are most predictive of Surgency, Negative Affectivity, and Orienting/Regulation. As previously noted, a first important future direction of this project is assessing the influence of prenatal depression, stress, and parental sensitivity on this association. A second future direction is to test the association of fetal movement and infant temperament constructs. A final possible future direction of this project is assessing how interventions to prevent prenatal and postpartum depression and stress influence the association between fetal markers and infant temperament. We believe that this project meaningfully contributes to our understanding of how infant biology, maternal characteristics, and environmental factors influence the development of temperament.

References

Als, H., Tronick, E., Lester, B. M., & Brazelton, T. B. (1977). The Brazelton neonatal behavioral assessment scale (BNBAS). *Journal of abnormal child psychology, 5*(3), 215-229.

Austin, M.-P., Hadzi-Pavlovic, D., Leader, L., Saint, K., & Parker, G. (2005). Maternal trait anxiety, depression and life event stress in pregnancy: relationships with infant temperament. *Early Human Development, 81*(2), 183-190.

Beauchaine, T. (2001). Vagal tone, development, and Gray's motivational theory: Toward an integrated model of autonomic nervous system functioning in psychopathology. *Development and Psychopathology, 13*(2), 183-214.

Beck, C. T. (1996). A meta-analysis of the relationship between postpartum depression and infant temperament. *Nursing research, 45*(4), 225-230.

Britton, J. R. (2011). Infant temperament and maternal anxiety and depressed mood in the early postpartum period. *Women & Health, 51*(1), 55-71.

Buitelaar, J. K., Huizink, A. C., Mulder, E. J., de Medina, P. G., & Visser, G. H. (2003). Prenatal stress and cognitive development and temperament in infants. *Neurobiol Aging, 24 Suppl 1*, S53-60; discussion S67-58.

Cohen, J. (1992). Statistical power analysis. *Current Directions in Psychological Science, 1*(3), 98-101.

Cohen, S., Kamarck, T., & Mermelstein, R. (1994). Perceived stress scale. *Measuring stress: A guide for health and social scientists*, 235-283.

Conger, R. D., Conger, K. J., & Martin, M. J. (2010). Socioeconomic status, family processes, and individual development. *Journal of Marriage and Family, 72*(3), 685-704.

Corporation, I. (2016). SPSS for Windows, version 24. In: IBM Corp Armonk, NY.

Cox, J. L., Holden, J. M., & Sagovsky, R. (1987). Detection of postnatal depression: development of the 10-item Edinburgh Postnatal Depression Scale. *The British Journal of Psychiatry, 150*(6), 782-786.

Derryberry, D., & Rothbart, M. K. (1988). Arousal, affect, and attention as components of temperament. *Journal of personality and social psychology, 55*(6), 958.

DiPietro, J. A., Costigan, K. A., & Voegtline, K. M. (2015). Studies in fetal behavior: revisited, renewed, and reimagined. *Monographs of the Society for Research in Child Development, 80*(3), vii.

DiPietro, J. A., Ghera, M. M., & Costigan, K. A. (2008). Prenatal origins of temperamental reactivity in early infancy. *Early Hum Dev, 84*(9), 569-575. doi:10.1016/j.earlhumdev.2008.01.004

DiPietro, J. A., Hodgson, D. M., Costigan, K. A., Hilton, S. C., & Johnson, T. R. (1996). Development of fetal movement--fetal heart rate coupling from 20 weeks through term. *Early Hum Dev, 44*(2), 139-151.

DiPietro, J. A., Hodgson, D. M., Costigan, K. A., & Johnson, T. R. (1996). Fetal antecedents of infant temperament. *Child Dev, 67*(5), 2568-2583.

DiPietro, J. A., Voegtline, K. M., Pater, H. A., & Costigan, K. A. (2018). Predicting child temperament and behavior from the fetus. *Development and Psychopathology, 30*(3), 855-870.

Doyle, C., Werner, E., Feng, T., Lee, S., Altemus, M., Isler, J. R., & Monk, C. (2015). Pregnancy distress gets under fetal skin: Maternal ambulatory assessment & sex differences in prenatal development. *Developmental Psychobiology, 57*(5), 607-625.

Gartstein, M. A., & Rothbart, M. K. (2003). Studying infant temperament via the revised infant behavior questionnaire. *Infant Behavior and Development, 26*(1), 64-86.

Gentile, S., & Galbally, M. (2011). Prenatal exposure to antidepressant medications and neurodevelopmental outcomes: a systematic review. *Journal of Affective Disorders, 128*(1-2), 1-9.

Gustafsson, H. C., Goodman, S. H., Feng, T., Choi, J., Lee, S., Newport, D. J., . . . Monk, C. (2018). Major depressive disorder during pregnancy: Psychiatric medications have minimal effects on the fetus and infant yet development is compromised. *Development and Psychopathology, 30*(3), 773-785.

Kagan, J., Reznick, J. S., & Snidman, N. (1988). Biological bases of childhood shyness. *Science, 240*(4849), 167-171.

Kop, W. J., Synowski, S. J., Newell, M. E., Schmidt, L. A., Waldstein, S. R., & Fox, N. A. (2011). Autonomic nervous system reactivity to positive and negative mood induction: The role of acute psychological responses and frontal electrocortical activity. *Biological Psychology, 86*(3), 230-238.

Kreibig, S. D. (2010). Autonomic nervous system activity in emotion: A review. *Biological Psychology, 84*(3), 394-421.

Lantz, P. M., House, J. S., Mero, R. P., & Williams, D. R. (2005). Stress, life events, and socioeconomic disparities in health: results from the Americans' Changing Lives Study. *Journal of health and social behavior, 46*(3), 274-288.

Liu, C. H., Giallo, R., Doan, S. N., Seidman, L. J., & Tronick, E. (2016). Racial and Ethnic Differences in Prenatal Life Stress and Postpartum Depression Symptoms. *Archives of Psychiatric Nursing, 30*(1), 7-12. doi:https://doi.org/10.1016/j.apnu.2015.11.002

Matthey, S., Henshaw, C., Elliott, S., & Barnett, B. (2006). Variability in use of cut-off scores and formats on the Edinburgh Postnatal Depression Scale: implications for clinical and research practice. *Arch Womens Ment Health, 9*(6), 309-315. doi:10.1007/s00737-006-0152-x

Mayberry, L. J., & Affonso, D. D. (1993). Infant temperament and postpartum depression: a review. *Health care for women international, 14*(2), 201-211.

McGrath, J. M., Records, K., & Rice, M. (2008). Maternal depression and infant temperament characteristics. *Infant Behavior and Development, 31*(1), 71-80.

Mezzacappa, E., Tremblay, R. E., Kindlon, D., Saul, J. P., Arseneault, L., Seguin, J., . . . Earls, F. (1997). Anxiety, antisocial behavior, and heart rate regulation in adolescent males. *Journal of Child Psychology and Psychiatry, 38*(4), 457-469.

Monk, C., Fifer, W. P., Myers, M. M., Bagiella, E., Duong, J. K., Chen, I. S., . . . Altincatal, A. (2011). Effects of maternal breathing rate, psychiatric status, and cortisol on fetal heart rate. *Developmental Psychobiology, 53*(3), 221-233.

Monk, C., Sloan, R. P., Myers, M. M., Ellman, L., Werner, E., Jeon, J., . . . Fifer, W. P. (2004). Fetal heart rate reactivity differs by women's psychiatric status: an early marker for developmental risk? *J Am Acad Child Adolesc Psychiatry, 43*(3), 283-290. doi:10.1097/00004583-200403000-00009

Ortiz, J., & Raine, A. (2004). Heart rate level and antisocial behavior in children and adolescents: A meta-analysis. *Journal of the American Academy of Child & Adolescent Psychiatry, 43*(2), 154-162.

Parade, S. H., & Leerkes, E. M. (2008). The reliability and validity of the Infant Behavior Questionnaire-Revised. *Infant Behavior and Development, 31*(4), 637-646.

Porges, S. W., & Byrne, E. A. (1992). Research methods for measurement of heart rate and respiration. *Biol Psychol, 34*(2-3), 93-130.

Ramírez, M. T. G., & Hernández, R. L. (2007). Factor structure of the Perceived Stress Scale (PSS) in a sample from Mexico. *The Spanish journal of psychology, 10*(1), 199-206.

Reebye, P. N., Morison, S. J., Panikkar, H., Misri, S., & Grunau, R. E. (2002). Affect expression in prenatally psychotropic exposed and nonexposed mother–infant dyads. *Infant Mental Health Journal: Official Publication of The World Association for Infant Mental Health, 23*(4), 403-416.

Remor, E. (2006). Psychometric properties of a European Spanish version of the Perceived Stress Scale (PSS). *The Spanish journal of psychology, 9*(1), 86-93.

Rothbart, M. K. (1989). Biological processes in temperament. *Temperament in childhood*, 77-110.

Rothbart, M. K. (2007). Temperament, development, and personality. *Current Directions in Psychological Science, 16*(4), 207-212.

Rothbart, M. K., Ahadi, S. A., & Evans, D. E. (2000). Temperament and personality: origins and outcomes. *Journal of personality and social psychology, 78*(1), 122.

Rothbart, M. K., Derryberry, D., & Hershey, K. (2000). Stability of temperament in childhood: Laboratory infant assessment to parent report at seven years. *Temperament and personality development across the life span*, 85-119.

Rothwell, D. W., & Han, C. K. (2010). Exploring the relationship between assets and family stress among low-income families. *Family Relations, 59*(4), 396-407.

Sanson, A., Hemphill, S. A., & Smart, D. (2004). Connections between temperament and social development: A review. *Social Development, 13*(1), 142-170.

Shiner, R., & Caspi, A. (2003). Personality differences in childhood and adolescence: Measurement, development, and consequences. *Journal of Child Psychology and Psychiatry, 44*(1), 2-32.

Siqueira Reis, R., Ferreira Hino, A. A., & Romélio Rodriguez Añez, C. (2010). Perceived stress scale: reliability and validity study in Brazil. *Journal of health psychology, 15*(1), 107-114.

Snidman, N., Kagan, J., Riordan, L., & Shannon, D. C. (1995). Cardiac function and behavioral reactivity during infancy. *Psychophysiology, 32*(3), 199-207.

Stifter, C. A., & Dollar, J. M. (2016). Temperament and developmental psychopathology. *Developmental psychopathology*, 1-62.

Stifter, C. A., Dollar, J. M., & Cipriano, E. A. (2011). Temperament and emotion regulation: the role of autonomic nervous system reactivity. *Developmental Psychobiology, 53*(3), 266-279.

Thompson, R. A., Lewis, M. D., & Calkins, S. D. (2008). Reassessing emotion regulation. *Child Development Perspectives, 2*(3), 124-131.

Werner, E. A., Myers, M. M., Fifer, W. P., Cheng, B., Fang, Y., Allen, R., & Monk, C. (2007). Prenatal predictors of infant temperament. *Developmental Psychobiology, 49*(5), 474-484. doi:10.1002/dev.20232

Zentner, M., & Bates, J. E. (2008). Child temperament: An integrative review of concepts, research programs, and measures. *International Journal of Developmental Science, 2*(1-2), 7-37.

Table 1

Demographic Characteristics and Comparisons

Variable	Emory	Emory *n*	Columbia	Columbia *n*	Comparison (ES)
Maternal Characteristics					
Age (*M*)	33.5 (4.32)	160	30.8 (6.66)	78	*d* = .49***
Caucasian (%)	91	160	37	78	--
Latina (%)	2	160	59	78	--
Married (%)	88	141	50	78	--
SES Characteristics					
< Bachelor's Degree (%)	19	160	44	78	*d* = .44**
Hollingshead (*M*)	51.36 (8.0)	160	--	--	--
Medicaid (%)	--	--	43	75	--
Pregnancy Characteristics					
Gravidity (*M*)	2.21 (1.39)	160	2.21 (2.38)	77	*d* = 0
Parity (*M*)	.74 (.89)	160	.99 (1.13)	77	*d* = .25
GA at birth (*M*)	38.52 (1.72)	160	39.35 (1.19)	78	*d* = .56***
Male Infants (%)	53	160	47	78	*V* = .05
Psychological Characteristics					
EPDS ≥ 10 (%)	55	155	22	46	*d* = .64***
EPDS ≥ 15 (%)	29	155	5	46	*d* = .64***
PSS (*M*)	29.74 (8.70)	156	25.94 (7.43)	78	*d* = .59***

*. Difference Significant at the 0.05 level.

**. Difference Significant at the 0.01 level.

Table 2

Fetal Markers Descriptive Statistics

	FHR			FHRV			Coupling		
	Emory $n = 160$	Columbia $n = 78$	Pooled $n = 238$	Emory $n = 160$	Columbia $n = 78$	Pooled $n = 238$	Emory $n = 160$	Columbia $n = 78$	Pooled $n = 238$
Mean	141.34	139.92	140.71	7.94	7.62	7.84	0.62*	0.65*	0.63
SD	10.57	7.47	9.98	4.31	2.78	3.86	0.11	0.09	0.10
Min	103.90	114.76	100.04	1.70	3.03	1.70	0.35	0.44	0.35
Max	168.76	158.90	168.76	31.06	17.14	31.06	0.90	0.89	0.90
Range	64.86	44.14	68.72	29.36	14.11	29.36	0.56	0.46	0.56

*. Difference is significant at the 0.05 level (2-tailed).

Table 3

Sex Differences: Fetal and Infant Measures

	FHR		FHRV		Coupling		SUR		NA		REG	
	M	F	M	F	M	F	M	F	M	F	M	F
	$n = 122$	$n = 116$	$n = 122$	$n = 116$	$n = 120$	$n = 113$	$n = 122$	$n = 116$	$n = 122$	$n = 116$	$n = 122$	$n = 116$
M	140.10	141.34	8.07	7.61	.64	.62	4.15	4.15	3.31	3.29	4.87	4.83
SD	10.92	8.88	4.31	3.32	.10	.10	.74	.76	.51	.47	.65	.73

Table 4

Infant Temperament Factors Descriptive Statistics

	SUR			NA			REG		
	Emory *n* = 160	Columbia *n* = 78	Pooled *n* = 238	Emory *n* = 160	Columbia *n* = 78	Pooled *n* = 238	Emory *n* = 160	Columbia *n* = 78	Pooled *n* = 238
Mean	4.09	4.27	4.15	3.37**	3.16**	3.30	5.09***	4.36***	4.85
SD	0.70	0.83	0.75	0.42	0.59	0.49	0.57	0.65	0.69
Min	2.51	1.66	1.66	2.58	2.08	2.08	3.67	3.20	3.20
Max	5.75	6.31	6.31	4.61	5.81	5.81	6.36	6.13	6.36
Range	3.24	4.65	4.65	2.03	3.73	3.73	2.69	2.93	3.16

**. Correlation is significant at the 0.01 level (2-tailed).

***. Correlation is significant at the 0.001 level (2-tailed).

Table 5

Data Correlation Matrices

Study origin		FHR	FHRV	Coupling	Surgency	NA	REG	GA fetal	GA birth
Emory	FHR	1							
	FHRV	-.02	1						
	Coupling	.12	.05	1					
	Surgency	-.01	-.04	.07	1				
	Negative Affectivity	-.03	.02	-.06	.02	1			
	Orienting/Regulation	.17*	-.04	.04	.52**	-.18*	1		
	GA fetal	-.24**	.11	.01	-.10	-.12	-.03	1	
	GA birth	-.23**	.02	.004	-.02	-.01	.11	.51**	1
Columbia	FHR	1							
	FHRV	.13	1						
	Coupling	-.02	-.03	1					
	Surgency	.09	-.21†	.24*	1				
	Negative Affectivity	.29*	.08	.04	.31**	1			
	Orienting/Regulation	-.06	-.21†	.29*	.59**	.13	1		
	GA fetal	.04	.09	-.11	.02	.22†	-.01	1	
	GA birth	-.14	.03	-.15	-.09	.06	-.18	-.05	1

*. Difference Significant at the 0.05 level.
**. Difference Significant at the 0.01 level.
†. .05 $p < .1$

Table 6

Surgency: Full/Final Model (Emory)

					Change Statistics				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	F Change	df1	df2	Sig. F Change
1	.11[a]	.01	.00	.70	.01	.88	2.00	157.00	.42
2	.13[b]	.02	-.01	.71	.01	.37	3.00	154.00	.78

a. Predictors: (Constant), GA birth, GA fetal session

b. Predictors: (Constant), GA birth, GA fetal session, Coupling, FHRV, FHR

Table 7

Surgency: Full Model (Columbia)

					Change Statistics				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	F Change	df1	df2	Sig. F Change
1	.16[a]	.03	-.01	.87	.03	.77	2	58	.47
2	.37[b]	.14	.06	.85	.11	2.44	3	55	.07

a. Predictors: (Constant), GA birth, GA fetal session

b. Predictors: (Constant), GA birth, GA fetal session, Coupling, FHRV, FHR

Table 8

Surgency: Final Model (Columbia)

					Change Statistics				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	F Change	df1	df2	Sig. F Change
1	.31[a]	.10	.07	.81	.10	3.80	2	70	.03

a. Predictors: (Constant), Coupling, FHRV

Table 9

Surgency: Full Model (Pooled)

					Change Statistics				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	F Change	df1	df2	Sig. F Change
1	.08[a]	.01	.00	.75	.01	.63	2	218	.54
2	.17[b]	.03	.01	.75	.02	1.73	3	215	.16

a. Predictors: (Constant), GA birth, GA fetal session

b. Predictors: (Constant), GA birth, GA fetal session, Coupling, FHRV, FHR

Table 10

Surgency: Final Model (Pooled)

					Change Statistics				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	F Change	df1	df2	Sig. F Change
1	.14[a]	.02	.01	.74	.02	4.49	1	231	.04

a. Predictors: (Constant), Coupling

Table 11

Negative Affectivity: Full/Final Model (Emory)

					Change Statistics				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	F Change	df1	df2	Sig. F Change
1	.13[a]	.02	.00	.42	.02	1.29	2	157	.28
2	.15[b]	.02	-.01	.42	.01	.38	3	154	.77

a. Predictors: (Constant), GA fetal session, GA birth

b. Predictors: (Constant), GA fetal session, GA birth, Coupling, FHRV, FHR

Table 12

Negative Affectivity: Full Model (Columbia)

					Change Statistics				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	F Change	df1	df2	Sig. F Change
1	.25[a]	.06	.03	.60	.06	2.01	2	58	.14
2	.37[b]	.13	.06	.59	.07	1.48	3	55	.23

a. Predictors: (Constant), GA fetal session, GA birth

b. Predictors: (Constant), GA fetal session, GA birth, Coupling, FHRV, FHR

Table 13

Negative Affectivity: Final Model (Columbia)

					Change Statistics				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	F Change	df1	df2	Sig. F Change
1	.29[a]	.08	.07	.57	.08	6.89	1	76	.01

a. Predictors: (Constant), FHR

Table 14

Orienting/Regulation: Full Model (Emory)

					Change Statistics				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	F Change	df1	df2	Sig. F Change
1	.15[a]	.02	.01	.57	.02	1.82	2	157	.16
2	.24[b]	.06	.03	.56	.04	1.96	3	154	.12

a. Predictors: (Constant), GA fetal session, GA birth

b. Predictors: (Constant), GA fetal session, GA birth, Coupling, FHRV, FHR

Table 15

Orienting/Regulation: Final Model (Emory)

					Change Statistics				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	F Change	df1	df2	Sig. F Change
1	.17[a]	.03	.02	.56	.03	4.58	1	158	.03

a. Predictors: (Constant), FHR

Table 16

Orienting/Regulation: Full Model (Columbia)

					Change Statistics				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	F Change	df1	df2	Sig. F Change
1	.21[a]	.04	.01	.64	.04	1.34	2	58	.27
2	.42[b]	.17	.10	.61	.13	2.85	3	55	.05

a. Predictors: (Constant), GA fetal session, GA birth

b. Predictors: (Constant), GA fetal session, GA birth, Coupling, FHRV, FHR

Table 17

Orienting/Regulation: Final Model (Columbia)

					Change Statistics				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	F Change	df1	df2	Sig. F Change
1	.36[a]	.13	.10	.61	.13	5.05	2	70	.009

a. Predictors: (Constant), Coupling, FHRV

Appendix A.

T-Test for Baseline vs. Post-Stressor Fetal Variables

		FHR baseline and rest		FHRV baseline and rest		Fetal coupling baseline and rest	
		= var. assumed	= var. not assumed	= var. assumed	= var. not assumed	= var. assumed	= var. not assumed
Levene's Test for Equality of Variances	F	5.87		.11		.63	
	Sig.	.02		.74		.43	
t-test for Equality of Means	t	1.27	1.26	-1.00	-1.00	-.37	-.37
	df	462.00	440.75	462.00	454.54	429.00	406.29
	Sig. (2-tailed)	.21	.21	.32	.32	.71	.71

Comparison of Fetal Values by Session Period

	FHR		FHRV		Coupling	
	Baseline	Post-Stressor	Baseline	Post-Stressor	Baseline	Post-Stressor
	n = 238	*n* = 226	*n* = 238	*n* = 226	*n* = 223	*n* = 198
Mean	140.71	139.42	7.84	8.21	.63	.63
Std. Deviation	9.98	11.83	3.86	4.17	.10	.11
Std. Error Mean	.65	.79	.25	.28	.01	.01

Surgency: Full Model Post-Stressor Fetal Variables (Pooled)

				Change Statistics				
Model	R	R Square	Std. Error of the Estimate	R Square Change	F Change	df1	df2	Sig. F Change
1	.11[a]	.01	.70	.01	1.10	2	189	.34
2	.14[b]	.02	.70	.008	.52	3	186	.67

a. Predictors: (Constant), Gestational age at birth, Gestational age at time of fetal session

b. Predictors: (Constant), Gestational age at birth, Gestational age at time of fetal session, FHR, FHRV, Coupling

Surgency: Post-Stressor Fetal Variables (Emory)

				Change Statistics				
Model	R	R Square	Std. Error of the Estimate	R Square Change	F Change	df1	df2	Sig. F Change
1	.11[a]	.01	.70	.01	1.002	2	154	.37
2	.14[b]	.02	.71	.007	.36	3	151	.78

a. Predictors: (Constant), Gestational age at birth, Gestational age at time of fetal session

b. Predictors: (Constant), Gestational age at birth, Gestational age at time of fetal session, FHR, FHRV, Coupling

Surgency: Post-Stressor Fetal Variables (Columbia)

				Change Statistics				
Model	R	R Square	Std. Error of the Estimate	R Square Change	F Change	df1	df2	Sig. F Change
1	.36[a]	.13	.67	.13	2.39	2	32	.11
2	.40[b]	.16	.69	.03	.33	3	29	.81

a. Predictors: (Constant), Gestational age at birth, Gestational age at time of fetal session

b. Predictors: (Constant), Gestational age at birth, Gestational age at time of fetal session, FHR, FHRV, Coupling

Negative Affectivity: Post-Stressor Fetal Variables (Emory)

				Change Statistics				
Model	R	R Square	Std. Error of the Estimate	R Square Change	F Change	df1	df2	Sig. F Change
1	.12[a]	.02	.42	.02	1.14	2	154	.32
2	.20[b]	.04	.42	.03	1.43	3	151	.24

a. Predictors: (Constant), Gestational age at birth, Gestational age at time of fetal session

b. Predictors: (Constant), Gestational age at birth, Gestational age at time of fetal session, FHR, FHRV, Coupling

Negative Affectivity: Post-Stressor Fetal Variables (Columbia)

				Change Statistics				
Model	R	R Square	Std. Error of the Estimate	R Square Change	F Change	df1	df2	Sig. F Change
1	.17[a]	.03	.42	.03	.45	2	32	.64
2	.19[b]	.04	.44	.01	.10	3	29	.96

a. Predictors: (Constant), Gestational age at birth, Gestational age at time of fetal session

b. Predictors: (Constant), Gestational age at birth, Gestational age at time of fetal session, FHR, FHRV, Coupling

Regulation/Orienting: Post-Stressor Fetal Variables (Emory)

				Change Statistics				
Model	R	R Square	Std. Error of the Estimate	R Square Change	F Change	df1	df2	Sig. F Change
1	.14[a]	.02	.57	.02	1.60	2	154	.21
2	.24[b]	.06	.56	.04	2.11	3	151	.10

a. Predictors: (Constant), Gestational age at birth, Gestational age at time of fetal session

b. Predictors: (Constant), Gestational age at birth, Gestational age at time of fetal session, FHR, FHRV, Coupling

Regulation/Orienting: Post-Stressor Fetal Variables (Columbia)

				Change Statistics				
Model	R	R Square	Std. Error of the Estimate	R Square Change	F Change	df1	df2	Sig. F Change
1	.41[a]	.17	.63	.17	3.29	2	32	.05
2	.53[b]	.28	.62	.11	1.52	3	29	.23

a. Predictors: (Constant), Gestational age at birth, Gestational age at time of fetal session

b. Predictors: (Constant), Gestational age at birth, Gestational age at time of fetal session, FHR, FHRV, Coupling