DOI: 10.1111/infa.12338

THEMATIC COLLECTION

INFANCY THE OFFICIAL JOURNAL OF THE INTERNATIONAL CONGRESS OF INFANT STUDIES WILEY

Individual differences in infancy research: Letting the baby stand out from the crowd

Koraly Pérez-Edgar[1] | **Alicia Vallorani**[1] | **Kristin A. Buss**[1] | **Vanessa LoBue**[2]

[1]The Pennsylvania State University, University Park, PA, USA

[2]Rutgers University, Newark, NJ, USA

Correspondence
Koraly Pérez-Edgar, 270 Moore Building, University Park, PA 16802, USA.
Email: kxp24@psu.edu

Funding information
This research was supported by National Institute of Mental Health Grants R01 MH109692 to Koraly Pérez-Edgar, Kristin Buss, and Vanessa LoBue, R21 MH103627 to Koraly Pérez-Edgar, and F31 MH121035 to Alicia Vallorani.

Abstract

Within the developmental literature, there is an often unspoken tension between studies that aim to capture broad scale, fairly universal nomothetic traits, and studies that focus on mechanisms and trajectories that are idiographic and bounded to some extent by systematic individual differences. The suitability of these approaches varies as a function of the specific research interests at hand. Although the approaches are interdependent, they have often proceeded as parallel research traditions. The current review notes some of the historical and empirical bases for this divide and suggests that each tradition would benefit from incorporating both methodological approaches to iteratively examine universal (nomothetic) phenomena and the individual differences (idiographic) factors that lead to variation in development. This work may help isolate underlying causal mechanisms, better understand current functioning, and predict long-term developmental consequences. In doing so, we also highlight empirical and structural issues that need to be addressed to support this integration.

1 | INTRODUCTION

We are all *different*, but when does a difference become an *individual* difference? That is, we are all aware of observable variation in behaviors and traits across the individuals we encounter in our daily lives. Building on this variation, scientists have worked to extract an understanding of systematic

and rule-bound functions that underscore broadly shared, and fairly universal, patterns of development. This nomothetic approach has advanced our understanding of core developmental functions, including language acquisition, motor behavior, and social cognition. For example, statistical learning approaches have presented infants with patterned input without overt markers of meaning or stimulus boundaries. By manipulating the inter-relations in the stimuli presented, we can track how infants extract units of analysis and meaning from streams of auditory input (e.g., transitional probabilities) to acquire language (Romberg & Saffran, 2010). Similarly, studies have revealed how infants identify and distinguish among faces (Nelson, 2001), build an understanding of numerical concepts (Xu, 2003), and come to appreciate the physical laws that govern the world around them (Carey & Spelke, 1994).

Within the confines of neurotypical development, these core sociocognitive mechanisms show ordered developmental trajectories and specific developmental mechanisms. These building blocks come together across development to create the more complex and varied phenomena that emerge over time. This includes subtle variation in language patterns across contexts, the ability to engage in contingent social interactions, and the acquisition of higher-order logical and critical thinking skills. As these building blocks come together, and the emergent products become more complex, researchers often report greater variation across participants and less predictive power when assessing performance (Lewis, 2000; Van Geert, 1998). Although some variation may be attributable to noise (e.g., method error), deviations may also be attributable to systematic differences in underlying mechanistic processes. The current paper examines how an individual differences approach can be used iteratively with a more universal approach, building on each other, to delineate and isolate developmental mechanisms and describe patterns of change.

For the purpose of this review, we consider three central facets in defining an individual difference. First, individual differences represent systematic variation in mechanisms that lead to an observed phenomenon of interest that can be predicted a priori. For example, children can differ broadly in their ability to process and identify faces. This variation is linked to differences in the tendency to focus on global or local signals in configural processing (Behrmann et al., 2006). Second, individual differences are reliably associated with an individual trait (e.g., temperament) or a specific context (e.g., socioeconomic status). To build on the current example, children on the autism spectrum are particularly challenged by face-processing tasks (Nomi & Uddin, 2015). Third, individual differences reliably predict developmental patterns and trajectories via the associated mechanisms. Again, while children on the autism spectrum are generally challenged in face-processing tasks, variation among these children can be linked with variation in configural processing and targeting face-processing mechanisms may improve functioning (Dawson, Webb, & McPartland, 2005).

An individual differences approach cannot replace a nomothetic approach, as they must actively inform each other in order to capture foundational patterns of development and identify when and where we see reliable deviations. However, understanding the form and function of individual differences can be important across many research traditions as it helps determine the reach, reliability, stability, and predictive value of developmental science. Indeed, a clear understanding of when, where, and for whom, mechanisms of interest are at play in development is a core feature, "necessary to our discipline's social utility" (McCall, 1977, p. 342).

In the next section, we discuss the foundational role of studies examining relatively universal (i.e., nomothetic) process in development. We then discuss when and how an individual differences approach can be used both to identify underlying mechanisms and to bolster predictive power for later functioning. Finally, we discuss examples of an integrative approach and how it might help researchers to overcome current barriers.

2 | NOMOTHETIC APPROACHES TO UNDERSTANDING DEVELOPMENT

Much of the infant literature has focused on basic cognitive and sociocognitive processes that can be captured in a controlled laboratory setting, with an eye to delineating typical patterns of development (Aslin & Fiser, 2005). The focus on fairly universal patterns of development with stable, shared mechanisms often allows for smaller, more homogeneous samples. The approach is to recruit infants, often clustered tightly around one or two age points, and examine performance focused on a specific concrete behavior across a number of fixed trials. For example, there are strong research traditions examining the emergence of object perception (Diamond, Cruttenden, & Neiderman, 1994) and face recognition (Nelson, 2001). From this line of work, we know that the ability to hold an unseen object in mind develops through the first two years of life, supported by prefrontal cortex development (Johnson, 2010), while early perceptual abilities shape an acquired ability to categorize and distinguish faces (Bornstein, Arterberry, & Mash, 2011).

When researchers actively examine for non-age-related differences among these studies, the focus is typically on patterns they themselves experimentally induced. For example, Bremner et al. (2005) examined the perception of object trajectories across occlusions. They systematically manipulated the width, distance, and time of occlusion comparing experimental and control conditions in each study. Overall, they found that 4-month-olds can perceive trajectory continuity only when the time or distance out of sight is short. Follow-up work in *Infancy* suggests that 4-month-olds can perceive horizontal and vertical trajectories, but not oblique trajectories, which may come on line by 6 months (Bremner, Slater, Mason, Spring, & Johnson, 2017). The authors suggest that this progression "likely relates to immature eye movement control" (p. 303), a potential (individual differences) mechanism that could be directly tested in follow-up studies.

The nomothetic approach has a number of strengths. First, the tight focus on a phenomenon often translates into experimental paradigms that isolate specific behaviors or processes. The paradigms can be implemented across laboratories, allowing for a more robust examination of development. Second, the focus on experimental controls and manipulations can translate into more straightforward statistical analyses that are easier to pre-register, conduct, report, and reproduce. Third, a focus on robust phenomenon can lead to more efficient and timely research as moderate sample sizes can still generate large effects. This is particularly beneficial since infant researchers work with a research population that can be hard to reach, cantankerous when recruited, and fairly indifferent to the researcher's need for clean crisp data. As Oakes (2017) points out, "few infant researchers feel awash with data" (p. 439).

The nomothetic approach is common in the recent infant literature. As an initial review, we characterized all of the papers in *Infancy* from January 2014 through December 2019. In that six-year period, the journal published 243 articles, presenting 307 separate experiments.[1] Of the relevant studies, 179 (60.5%; Figure 1) examined variation at the level of the group, based on either age or experimental manipulation. This pattern reflects general trends across other journals as well (Mills-Smith, Spangler, Panneton, & Fritz, 2015; Oakes, 2017).

The studies in this six-year period averaged 27.5 infants per cell with a median of 20 ($SD = 27.0$; Figures 1 and 2). On this point, *Infancy* does not seem to be atypical. Oakes (2017) reviewed 70 articles using infant looking-time measures published from 2013 to 2015 in nine leading developmental

[1]For the purpose of this discussion, we removed ten publications because they were review papers or used a non-human model. One additional paper involving a highly skewed sample sizes ($N = 117{,}881$) and was also removed. As a result, 296 studies were assessed for this manuscript.

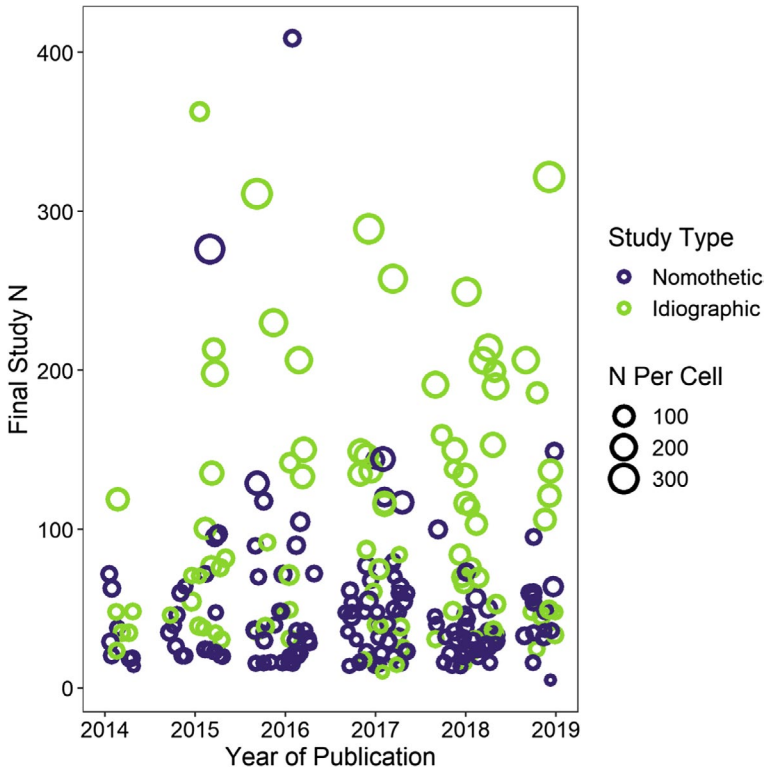

FIGURE 1 Number of infants per cell in studies published in *Infancy* from January 2014 through December 2019. Individual studies within a single publication are noted separately. Studies are characterized for having an individual difference component. Circle size indicates number of participants per cell

journals. She found that many published with 10–15 infants in the study or per cell. Of these, 11% had samples greater than 25 per cell. These numbers reflect study design protocols that are predicated on targeting fairly stable phenomena with relatively large effect sizes (DeBolt, Rhemtulla, & Oakes, 2020, under review; Eason, Hamlin, & Sommerville, 2017). Another feature of the nomothetic approach has been the general reliance on cross-sectional studies. For example, in the same 6-year window, 95.0% of the nomothetic papers were cross-sectional (Figure 3). This compares with 54.7% of the studies taking an individual differences approach.

Oakes (2017) suggests that even with seemingly robust effects, larger sample sizes may be needed. In her review, she presented a series of analyses drawing from her own data, starting with a sample size of over 30 in each of three independent studies. Sampling from each study, analyses were re-run with steadily decreasing cell sizes (24, 20, 16, 12, and 8 infants) with 1,000 random subsamples. The most fully powered comparisons found clear and robust group differences linked to the experimental manipulation. However, as the cell sizes diminished, the distinction became "blurry" and fewer and fewer results emerged.

Despite the evident strength of the nomothetic approach, there are limitations to the scope of questions that can be addressed. As noted earlier, as infants get older and are exposed to a wider range of experiences, our constructs of interest become more varied, multi-dimensional, and often embedded within personal or contextual processes. For example, phoneme detection is followed by receptive language, which is then followed by language production. At each step in the chain, there is a marked increase in the processes that can influence the construct of interest (e.g., language complexity; richness

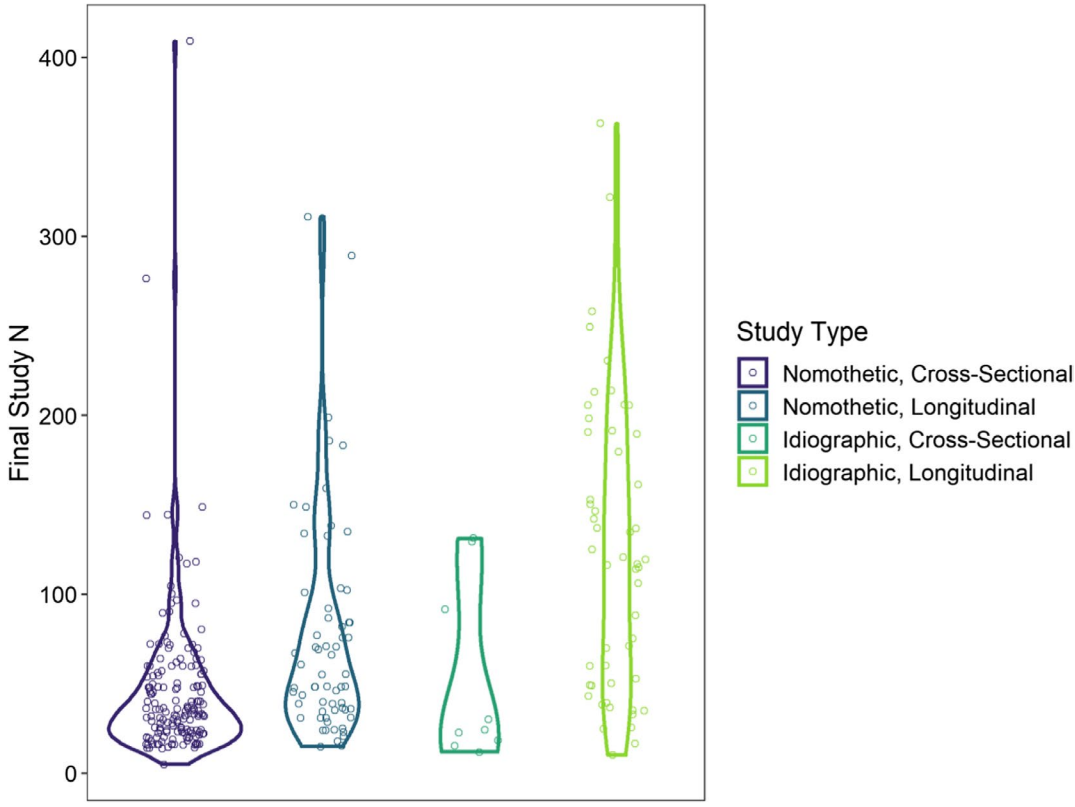

FIGURE 2 Violin plots illustrating distribution of studies in *Infancy* from January 2014 through December 2019. Studies are presented separately based on their classification as nomothetic or idiographic, as well as cross-sectional or longitudinal in order to depict the distribution of study sample sizes

of the linguistic environment), and the ways in which the construct can be expressed. This, in turn, increases the variance of observed data, leading to a slow decline in effect sizes and predictive power as our constructs of interest become more complex. This variance is often rule-bound and governed by systematic individual differences. Directly comparing candidate mechanisms that generate observed individual differences are important for both practical and conceptual reasons.

First, with increases in variance it can be more difficult to extract the true effect of the experimental manipulation or intervention. For example, there is continuing controversy regarding the short- and long-term benefits of intervention programs, such as early Head Start (Barnett, 2011). Well-designed randomized control trials are sometimes stymied by reporting small to null effects sizes. On its face, the interpretation is that the manipulation failed to move the target of engagement. However, it is often the case that some participants show large and meaningful improvement, while others show no change. Thus, systematic variations in intervention impact are hidden when we collapse across the entire sample. Understanding associated individual differences can reveal when, and for whom, an intervention has an effect.

Second, at a conceptual level, isolating patterns of change can reveal specific mechanisms that help refine and test our theoretical models. This is a gap in the literature since portions of the published research take pains to generate careful and precise descriptions of a phenomenon, but do not move to the next step of outlining potential causal mechanisms. One potential step is to use computational models which can generate plausible mechanisms underlying observed infant data, tweaking the model to best

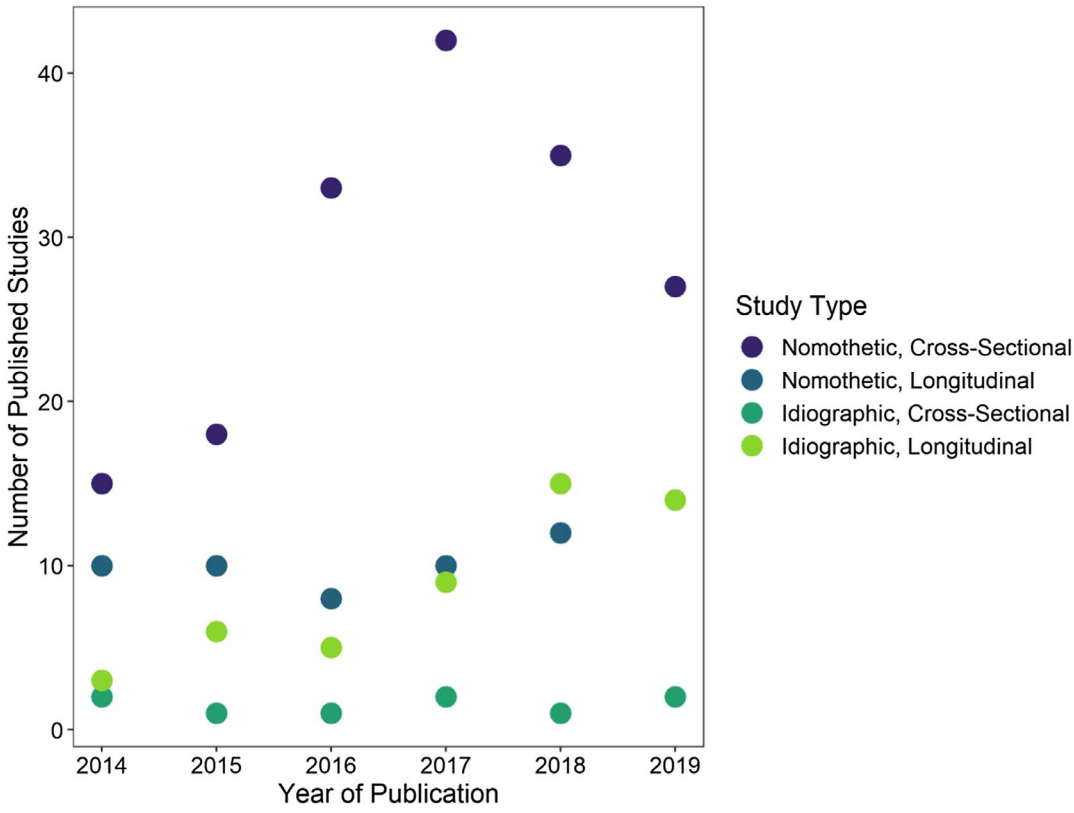

FIGURE 3 Characterization of published studies in *Infancy* from January 2014 through December 2019. Individual studies within a single publication are noted separately. Studies are characterized as nomothetic or idiographic, as well as either cross-sectional or longitudinal

approach phenotypic patterns (Mareschal, 2000; Mareschal & French, 2000). These mechanisms can then be tested through experimental manipulations, or naturally occurring individual variation in the mechanism of interest (Allman & Mareschal, 2016).

3 | IDIOGRAPHIC APPROACHES TO UNDERSTANDING DEVELOPMENT

One approach to probing for mechanism is to experimentally manipulate presumed mechanisms. These data can reveal the plausibility of a purported mechanism—that is, the mechanism *can* impact the outcome of interest (Pérez-Edgar & Hastings, 2018). The second approach looks to "naturally" occurring variation in a putative mechanism to see whether the variation in mechanism tracks variation in outcome and therefore actually *does* impact the outcome of interest. Although both approaches are interdependent, iteratively building on data, they are often carried out in parallel, dominating different subfields within developmental science.

The observational approach is often evident in individual differences research focused on socioemotional development. The questions of interest are often correlational, focusing on individual traits that the child carries with them into the lab. These questions include gender differences in emotional expression (Chaplin & Aldao, 2013; Chaplin, Cole, & Zahn-Waxler, 2005) and temperamental risk for

anxiety (Fox, Snidman, Haas, Degnan, & Kagan, 2015; Kagan, 2018a, 2018b). A parallel approach focuses on factors characterizing the context infants and children are embedded in. This includes the impact of maternal depression on attachment (Martins & Gaffan, 2000) and the impact of socioeconomic status on emotion regulation (Noble, Houston, Kan, & Sowell, 2012). These studies take an observational approach, reflecting the reality that the forces at play cannot be randomized or are challenging to manipulate.

The observational approach often taken in idiographic studies also reflects historical forces that worked to push experimental and observational studies further apart. In the 1970s, many leaders of the field fretted that we lacked a "substantial science of naturalistic developmental processes" (McCall, 1977). Wohlwill (1973) argued that developmental psychology would devolve into a paler branch of general psychology defined simply by the age of the participants if it did not focus on *development*, or, rather, change over time. Work focused on development was "to remain at an essentially descriptive level" (Wohlwill, 1973).

The consequences of this movement are still evident today. For example, in our review of six years of *Infancy,* 39% of studies incorporated an individual differences approach. Yet, these studies account for 85% of the longitudinal studies published in the same time window. The division between nomothetic and idiographic approaches is such that individual differences researchers sometimes eschew experimental manipulations, even when they may be crucial for isolating mechanisms of interest.

There are, of course, a number of challenges to carrying out studies focused on individual differences. First, this approach requires larger sample sizes than typically needed of group-based nomothetic studies. This reflects an increase in measure variance and a yoked decrease in effect sizes. Recent reviews have criticized the small sample sizes in experimental infant work (DeBolt et al., 2020, under review; Frank et al., 2017; Oakes, 2017). In our review of *Infancy*, the individual differences papers had an average total sample size of 108.92 (range from 10 to 1,459) and an average of 92.41 participants per cell (range from 5 to 1,459). While the averages are certainly larger than seen with the nomothetic papers, the range suggests that concerns with power and robust measures are likely central in this literature as well.

Many of the barriers that lead to small sample studies at one time point—recruitment, testing, staffing—are also in play when contemplating increasing the sample size or adding the additional layer of retention for longitudinal work. Kenny and Judd (2019) suggest that "power analyses always rest on a series of informed guesses" (p. 7). As with nomothetic research, studies taking an individual difference approach also tend to be highly optimistic when predicting power and needed sample sizes. Mills-Smith et al. (2015) coded 158 papers (from 2007 through 2012) from six leading journals in developmental psychology for core characteristics of their demographics, design, and provided statistical outputs. They noted a positivity bias inflating the size of potential effects—particularly in large samples. As such, it may be that infant measures are not as robust as they seem. Thus, less optimistic power assessments are needed, building in an added "cushion" of recruiting more than otherwise indicated when reading the literature or carrying out standard power projections (Faul, Erdfelder, Lang, & Buchner, 2007; Muthén & Muthén, 2002).

As one example of a longitudinal and idiographic approach, Izard, Hembree, and Huebner (1987) examined differential emotions theory in young infants, with an interest in seeing if developmental change in emotional expression is orderly over time. Infants were observed while receiving inoculations from the pediatrician at 2, 4, 6, and 18 months. Signs of pain decreased over time as signs of anger increased. Thus, there appears to be a shift in expressed emotion with development, even in the face of an identical trigger. In addition, the shift was systemic at the level of the group over time.

However, not all infants showed the same amount of pain or anger across testing points. Rather, there was a wide range of expression, from the mild whimper to howls of displeasure. This variation was also not random in that the authors noted stability in the rank ordering of the infants over time.

The focus on phenotypic change over time can help researchers capture both average trajectories and the individual's place within that trajectory. However, it is important to note that this study (Izard et al., 1987) was only modestly powered with a total sample size of 25.

Beyond recruiting and retaining a large sample of infants, there are additional difficulties facing individual differences research. Often, the moderators and variables of interest are highly complex and difficult to capture. As a result, researchers sometimes cannot assess the full observable range of a measure as needed for a well-powered analysis. For example, studies examining the link between parental behavior and infant socioemotional outcomes often capture only one portion of the behavioral spectrum, restricting the available relations that may emerge (Lamb, 2015).

Although not an infant study, work by Deater–Deckard Dodge Bates and Pettit (1998) illustrates the promise and complication of this approach. They examined 20 risk variables in a large sample of 566 children, measuring externalizing problems annually from ages 5 to 10. The 20 variables were grouped into four categories of risk (child, sociocultural, parenting, and peer-related). The large sample size, and the extensive multi-method assessment, allowed researchers to examine two forms of individual differences. First, they looked to see whether they could detect stability in risk factors within the high-risk group relative to the comparison group. Second, they looked to see whether the predictive value of the four categories of risk variables shifted across time. Thus, they could examine variation and stability both within and across groups.

Even when large samples and multiple measures are available, idiographic studies are still vulnerable to a general concern that the individual differences literature often focuses on markers or traits that the child carries with them to a study and to the specific question of interest or analysis (Scarr, 1992). While we may screen and recruit for one specific trait (e.g., temperament, parental diagnostic status), it is unlikely that the variable of interest exists in isolation relative to the myriad of other factors that may influence the outcome of interest. That is, individual factors are rarely orthogonal to each other and the variables or mechanisms of interest are neither randomly assigned nor randomly distributed.

In addition, traits of interest often defy "yes" or "no" designations. With questions of ethnicity, temperament, parental characteristics, and gender, there are often variations across a continuum, which complicates our ability to link a specific trait to a specific outcome. This ambiguity and lack of control pushes against one of the tenets of experimental research. Namely, the gold standard is to assign, isolate, manipulate, and then measure the target mechanism so that we can claim causality for any observed variation in outcome among participants (Goodhew & Edwards, 2019).

Ironically, another central concern with the idiographic approach is that individual differences studies do not always center individual differences in their design and analyses. Rather, studies often rely on groupclassifications as the central participant-level factor of interest. While observing group-linked differences is a beneficial first step in individual differences research, an understanding of underlying mechanisms is needed in order to help understand observed variation (Buss, Davis, Ram, & Coccia, 2018; Buss & Qu, 2018).

To illustrate, nationality is often used as a proxy for culture. The intention is to look for variation in an outcome as a function of a culturally linked behavior. While a comparison between nation A and nation B is meant to represent the presumed variation in cultural practices, analysis of an individual differences variable must assess the mechanisms in question, note the distribution of the mechanism, and see whether variation in the mechanism is associated with the observed outcome both within, and across, cultural groups (Carroll, 1978).

For example, cultural norms and ideals shape how we come to assess maternal sensitivity. Non-contingent, dismissing, and overly intrusive behaviors are linked to maladaptive socioemotional profiles, marked by increased negative affect and poor self-regulation skills (Kiel & Kalomiris, 2015). Cross-cultural work suggests that sensitivity is not necessarily tied to a single behavioral profile.

Rather, maternal sensitivity is defined by child outcomes, which, in turn, are mediated by the match between maternal behaviors and cultural expectations (Friedlmeier & Trommsdorff, 1999). In Germany, mothers focused on the cause of an emotion when helping the child regulate and meet the culturally approved target response. Mothers who scaffolded independent and instrumental responses were deemed more sensitive. In contrast, Japanese mothers often targeted the child's emotional display in response to an emotional elicitor. Here, sensitivity was embedded in the ability to mold emotion expressions in support of harmony within the social group. Thus, sensitive Japanese mothers focused on shaping and mirroring facial expressions. One could imagine that a uniform level of maternal sensitivity would not be conveyed where we to "swap" these specific social behaviors across contexts.

Both idiographic and nomothetic approaches bring clear strengths to our attempts to understand development. There are also clear concerns that must be bridged. From the nomothetic approach, there is the argument that individual differences are not relevant because researchers see very little variability in the measures of interest. However, many tasks have been specifically designed to minimize variability (Goodhew & Edwards, 2019). Thus, it is hard to detect what you have systematically eliminated. From the idiographic approach, there is the concern that more experimental designs will narrow the types of constructs that can be studied. However, greater methodological control, and the use of experimental manipulations, when possible, will allow for a more dynamic and nuanced view of individual differences, which are often treated as static and monolithic within an individual child.

4 | COUPLING NOMOTHETIC AND IDIOGRAPHIC APPROACHES

Merging two approaches allows researchers to examine naturally occurring variation in a trait or function and then see whether there is further variation in its expression with controlled changes in context, content, or motivation. In addition, researchers can see what factors disrupt, or change, the general developmental trajectory observed for a construct or skill of interest. Thus, we can better approach answers that will interest a larger swath of scientists than we often see now. This approach also provides insight into both interindividual differences and intraindividual change.

By widening the number of questions that researchers can ask, we may also remove some of the risk involved in charging forward with a complex study resting on the shoulders of young infants. There are four approaches a shared nomothetic-idiographic study can take, together or in isolation. First, we can capture change in a system of interest. Second, we can illustrate how variation in the form or rate of change can impact an outcome. Third, we can propose mechanisms that can predict variations in change. Fourth, we can assess the consistency of observed patterns with variation in the proposed mechanism and context. This approach can be illustrated across domains of early development both within and across studies.

For example, a rigorous series of studies has found that infants show a distinct attention bias to emotional faces at 7 months, that is not evident at 5 months (Peltola, Leppanen, Maki, & Hietanen, 2009; Peltola, Leppanen, Palokangas, & Hietanen, 2008; Peltola, Leppanen, Vogel-Farley, Hietanen, & Nelson, 2009). This pattern is thought to reflect the infant's new found ability to both distinguish variation in emotional expression and extract meaning from the expression, which may trigger selective attention biases (Burris et al., 2019; LoBue, Kim, & Delgado, 2019). These strongly characterized nomothetic data on attention bias set the foundation for recent work linking early face processing to the later presence of individual differences in attachment patterns (Peltola, Forssman, Puura, IJzendoorn, & Leppänen, 2015) and prosocial behavior (Peltola, Yrttiaho, & Leppänen, 2018).

A second example examined the progression of the A-not-B task mastery in a longitudinal sample from ages 6 to 12 months (MacNeill, Ram, Bell, Fox, & Pérez-Edgar, 2018). The time window

is particularly crucial, as it allows researchers to follow the entire life span of skill acquisition, from floor to ceiling. The authors found evidence for the theorized non-linear acquisition of object permanence (Wellman, Cross, & Bartsch, 1987) with logistic growth curve models, noting that infants with faster performance rates reached performance milestones earlier. In addition, infants with faster rates of increase in A-not-B performance had lower occipital EEG power at 6 months and greater linear increases in occipital EEG power over the course of the study. Thus, the study captured variation in change in underlying linear mechanisms associated with observed individual differences in non-linear skill acquisition, as often predicted by dynamic systems models (Van Geert, 1998).

As a third example, Kuchirko, Tafuro, and Tamis LeMonda (2018) examined 190 Mexican-American, Dominican-American, and African-American mother–infant pairs, assessed at ages 14 and 24 months. With their large sample, they set out to examine two separate, but integrated questions. First, as a group, how does contingency between the pairs change as the infant produces, and observes, more sophisticated gestures and responses over time? Second, at the individual level, do individual differences in responsivity among mothers associate with responsiveness to vocalizations and gestures by infants? Thus, the single study was able to address important questions in early development, including (a) the form and function of a core communicative tool, (b) change in the communicative tool over time at a crucial window in development, and (c) how variation in potentially associated mechanisms impacts the presence and change of the communicative tool. The authors were able to address multiple questions by basing the study design on known developmental milestones and evident theory and data on the processes of communication. In this work, the authors were able to show that this integrated network as "simultaneously culture-specific and culture-general" (p. 573).

As a final example, LoBue, Buss, Taber-Thomas, and Pérez-Edgar (2017) examined attention patterns to putative threats in a cross-sectional sample of infants ages 4–24 months of age. LoBue and colleagues found no significant age differences in non-social threat cues, suggesting that a perceptual bias for threat is presented early in life and stable across infancy. However, when presented with social cues, there were age-related changes in infants' responses. These data suggested that different developmental mechanisms may underscore attention biases to social and non-social stimuli. In addition, the age effects suggested greater opportunity for individual differences to emerge for social cues, as increased variation appeared to emerge over time. A follow-up study (Pérez-Edgar, Morales, et al., 2017), in turn, found that dwell time to social stimuli was associated with subsequent orienting. These findings suggest that although age (nomothetic) was directly associated with an emerging bias to threat, the *impact* of processing threat on subsequent orienting was associated with age and temperament (idiographic).

These examples illustrate the feasibility and power of bringing together both experimental and observational methods across a variety of ages and research questions. And yet, multiple surveys of the infant literature (Eason et al., 2017; Mills-Smith et al., 2015; Oakes, 2017) suggest that this is not a common approach. Thus, the final section of this review overviews some of the causes and responses to this traditional separation.

5 | CURRENT BARRIERS TO INTEGRATING NOMOTHETIC AND IDIOGRAPHIC APPROACHES

5.1 | Empirical barriers

Frank et al. (2017), in introducing the ManyBabies consortium, suggest that the costs of infant research are compounded by the investment of time and effort needed for recruitment and testing, leading to small samples with limited power. Frank and colleagues suggest that researchers are then

faced with two options. Do you standardize the protocol to eliminate variables or do you deliberately increase heterogeneity in order to study variability? These options are not always compatible. Frank et al. (2017) focused on identifying and directly examining heterogeneity in the variables of interest, as well as materials and experimental methods, rather than variability across individuals. However, the concerns carry over to individual differences studies.

As a comparison, Eason et al. (2017) note that the focus in adult studies is on eliminating between-participant variation in order to focus on within-person variation. Often, of course, this approach is predicated on the fact that the researcher is taking a within-subjects repeated measures approach. With a repeated measures design, researchers can boost power, even at a smaller sample size, by accounting for both within-person and between-person variability.

5.2 | Psychological barriers to focusing on individual differences

Our training as researchers often emphasizes the hunt for robust mechanisms and processes that have a meaningful, and widespread, impact on development (McCall, 1977; Wohlwill, 1973). Meaningful can have a variety of definitions, and often centers on the practical significance a mechanism may have on the daily functioning of children. While this standard may be difficult to quantify, researchers try, at the very least, to show that the mechanism and process have robust and reliable statistical significance (McCartney & Rosenthal, 2000). Researchers are rewarded scientifically—and professionally—if they can demonstrate effects that are strong, easily replicable, and fairly universal. These effects, by inference, are *important.* Thus, also by inference, effects that are qualified by "it is complicated," "it depends," and "sometimes," are considered less important, if not, *unimportant.*

Even in the face of limited time and resources, there are specific benefits of focusing on phenomena that may only hold true for smaller pockets of the population or under specific circumstances. First, although a specific mechanism or process may only hold for a subset of children, you can then examine and target mechanisms that will have an outsized effect on their developmental trajectory among those children. Indeed, this is almost the ideal pattern of findings for the idiographic approach, as we can then focus resources on infants at risk. For example, deficient phenylalanine hydroxylase impairs the metabolism of phenylalanine, negatively impacting neural and intellectual development (Guldberg et al., 1998). Our understanding of the mechanisms underlying phenylketonuria led to modified diets that minimized cognitive delays once thought unavoidable (Giżewska, 2015).

Second, a scientific tolerance for individual differences can help us refine our understanding of how and when mechanisms come into play—knowledge that can be applied at a broader level. Variation in infant motor development can illustrate this point. Typically, motor milestones are reached in a fairly prescribed order with a well-described developmental window in typically developing infants. Indeed, deviation from this trajectory is used as an early sign of developmental delay that can spill over into cognitive and social domains (Iverson, 2010; Leonard & Hill, 2014).

However, many of our assumptions concerning the orderly, programmed, and universal nature of motor development is derived from fairly homogenous samples of Western infants that share both common genetic backgrounds and daily life experiences. Recent cross-cultural work suggests that child-rearing practices can both accelerate and delay the acquisition of motor milestones, as well as shift the trajectories of acquisition (Karasik, Adolph, Tamis-LeMonda, & Bornstein, 2010). Importantly, common themes can be pulled from observed heterogeneity in patterns of motor development. For example, although infants may differ in when and how they engage in object manipulation, all infants rely on this process in order to learn about their environments (Adolph & Robinson, 2015).

Beyond an esthetic preference for robust effects, there is also an emerging concern with moderation effects within the context of psychology's current replication crisis (Bergmann et al., 2018; Frank et al., 2017). The suspicion is that studies that rely on moderators (in this case individual differences) reflect patterns of unreported *p*-hacking (Simmons, Nelson, & Simonsohn, 2011). Moderators are suspected to be an a-theoretical way of slicing the sample in order to produce at least one pocket of data that reaches the $p < .05$ threshold. And, indeed, there are cases in which moderators are introduced after the fact in order to help explain a puzzling (lack of) findings. This concern, however, conflates two important patterns of contributions to the research literature. First, there are strong literature traditions that identify specific individual differences factors that are tied to discrete theory, supported by extant data, that can be used to a priori motivate a study that is fully powered and systematic in assessing potential moderators. Second, exploratory analyzes, when properly labeled, can help bolster theory formation and help guide subsequent studies. Exploratory analyses then serve as the foundation for future, confirmatory, studies with expected variation embedded in the design.

5.3 | Structural impediments to individual differences research

Screening and recruiting participants in sufficiently large numbers for individual differences to emerge can be slow, labor intensive, and expensive (Frank et al., 2017). It is also work that must often be carried out collaboratively both within and across labs in order to reach recruitment goals. However, students and early career researchers often face incentive structures that disincentivize work that is slow, labor intensive, and team-based. A small laboratory may be able to recruit and test roughly 30 infants over the course of a year. However, it is likely unfeasible to expect that same laboratory to recruit five times as many participants—at least not in a timely manner.

For early career researchers, the clock begins the day they start their position. The fruits of their labor, quantified in publications, are expected to roll out in a timely and consistent manner (Duffy, Jadidian, Webster, & Sandell, 2011). How do you then assess collaborative work for early career researchers within this evaluation structure? The field may need to step back and explicitly value leadership and participation in collaborative work as an indicator of a researcher's contribution as an individual scientist. The individual researcher's expertise is now embedded in a larger series of measures and variables that are, in turn, partially overseen by another collaborator. Thus, metrics of authorship or intellectual ownership cannot rely on stand-alone pieces of evidence that perpetuate the image of the lone scientist singularly generating new knowledge.

Given the empirical, psychological, and structural barriers to some forms of individual differences work, there will have to be continued dialogue on how to more fully integrate the approach into infant research. Here, we highlight central questions or practices that should be addressed.

6 | POTENTIAL NEXT STEPS

6.1 | Incorporating repeated measures

Researchers are sometimes concerned that repeated measures designs may overtax infants. In response, many laboratories build in three common safeguards for infants involved in complex protocols. First, breaks are built in throughout the session, and of course, unplanned breaks are taken if infants are fussy, hungry, or need a diaper change. An important consideration, of course, is to work with the caregiver to schedule visits for a stretch of time during the day when the infant is likely to be

alert and even-tempered. Second, tasks varying in stimuli, level of arousal, or behavioral challenge are interwoven so that the infant is neither bored nor over-taxed. Third, visits can be split across multiple days. Often, we are reluctant to rely entirely on multiple visits as it increases the participation burden on families and increases the risk of missed visits and missing data. However, one strategy is to rank order measures or tasks by order of importance and front load those into the first visit—assuming of course that there are no empirical or theoretical concerns with order effects.

There are, of course, concerns that some laboratory designs are susceptible to order or bleed-over effects. Given the crucial importance of habituation (induced boredom) and novelty preference (elicited interest) in looking-time studies (Hunter & Ames, 1988; Rubio-Fernández, 2019), it may be that exposure to one exemplar may make it impossible to disentangle the response to the alternate exemplar in a repeated measures design. This danger is particularly acute if testing is completed in one laboratory session. It may be that infants will need to come back twice, at some set time lag, in order to employ a within-subjects design. Another approach is to use an accelerated longitudinal design in which infants are brought in at varying ages (e.g., 6, 12, 18 months) and then repeatedly tested at regular intervals (e.g., three times spaced every 6 months). In this way, researchers can span a large age range (6–30 months) in a shorter period of time. This approach can also help disentangle the impact of development (maturation) from experience (repeated testing). Importantly, the testing schedule must match the rate of change for the phenomenon of interest. For example, the noted A-not-B study required monthly testing (MacNeill et al., 2018), while the study of externalizing risk targeted change year-over-year (Deater–Deckard, 1998).

An alternative is a planned missingness design (Little & Rhemtulla, 2013), which allows researchers to build in a known pattern of missing data by randomly assigning an infant to skip a specific task or, in a longitudinal study, skip a specific testing time point. In this way, researchers can soften the logistical and recruitment burden of a fully complete design, while ensuring that any missing data are likely to be missing at random rather than systematic traits or conditions that favor data collection in one condition, but not the other. In either case, many of the steps taken by laboratories to build rapport and encourage study engagement across longer scale longitudinal studies can be used to minimize attrition with more tightly timed visits.

Another example embeds multiple conditions within a task or paradigm. A-B-A or A-B-C designs allow the researcher to chart individual and group changes in a systematic manner. For example, Buss and colleagues (Buss, 2011; Buss et al., 2018) had toddlers complete a set battery of social–emotional episodes designed to reflect low (e.g., puppet), medium (e.g., adult stranger), and high (e.g., mechanical spider) threat contexts. They found that all but the most exuberant toddlers showed signs of fear in the high threat environment. This is to be expected and would be an example of a low variability, fairly nomothetic response to threat.

However, they also looked to see how the same children reacted to the medium and low threat conditions. For most children, they observed an expected decrease in behavioral signs of stress as well as less stress reactivity. For a subset of the toddlers, fear and stress levels remained high, regardless of the ostensible decrease in threat. These toddlers displayed high levels of dysregulated fear, a temperamental trait associated with systematic patterns of functioning at biological, social, and cognitive levels, that is, also associated with an increased risk for anxiety in middle childhood and adolescence.

Here, variation is not ever-present or persistent—these children do not cower at all times from all things. However, a careful manipulation of theorized mechanisms allowed for individual differences to emerge and be studied. While one strategy is to have a single infant complete multiple variations of a specific condition or task, a parallel approach is to also have a single infant complete a battery of tasks that work to either (a) provide validation for a specific construct across multiple operationalizations or (b) provide insight into the mechanisms that help explain a central skill or trait of interest (see LoBue et al., 2020, for a more detailed review).

6.2 | Power and sample size

Once laboratory procedures and measures are agreed upon, researchers need to clarify how many infants should participate in the study. As noted, individual differences studies typically require larger samples in order to capture and assess any systematic patterns of variation. However, what does it mean to increase the sample size? The initial, and most straightforward, assumption is that researchers should recruit and test more participants. In studying the impact of temperamental negative affect on the later expression of anxiety, researchers should screen and recruit enough infants to represent the full array of this specific temperamental trait, increasing the odds that they will see a range of associated outcomes (Fox & Pine, 2012). Of course, this is easier said than done. Recent work has acknowledged the cost–benefit analysis of recruiting ever more babies for a single study. As such, Schott, Rhemtulla, and Byers-Heinlein (2019) suggested that principled and transparent protocols for assessing data both before and during data collection can help researchers determine whether they have tested sufficient numbers to generate robust and replicable findings.

However, sample size goes beyond the individual number of infants enrolled as researchers need to affirmatively designate the unit of analysis. Typically, this appears to be a straightforward question—individual infant are used as the unit of analysis. Multiple trials within a condition are often averaged to create a single score for the individual. In contrast, if researchers have multiple trials per infant, and enough infants in the sample, they can focus on individual trials as the unit of analysis. Doing so allows researchers to use every source of data available from an infant and capture a richer characterization of variation than can be extracted when collapsing across trial to create a single summary value per infant (Baayen, Davidson, & Bates, 2008).

This approach is common in some areas of psychology, taking advantage of robust testing phenomena. An example can be drawn from the (adult) visual attention research (Goodhew & Edwards, 2019). Here, effect sizes are so large that study protocols often call for an inordinate number of trials per condition (in the hundreds) from exceedingly small samples (e.g., single digits to low teens; Rademaker, Park, Sack, & Tong, 2018). As studies move up the visual perception ladder, the ratio of trials to participants begins to shift closer to patterns typically seen in other areas of research. Small samples coupled with many trials are less common in the infant literature.

Of course, item level analyses are not without pitfalls. Assuming that trials are independent units, as typically done with infants participating in the laboratory, may lead to potentially erroneous conclusions. For example, Gustafson, Sanborn, Lin, and Green (2017) examined variation in infant cries. Typically, parents can quickly learn to detect their infant's "cry signature" from among other infants. With training, non-parent young adults can learn to do the same. Prior work had suggested that infants also had a culturally linked signal to their cries, even in the first months of life. Gustafson Sanborn Lin and Green (2017) found that they could replicate culture-linked differences in cry characteristics (e.g., pitch and tone) when comparing infants from US English-speaking homes and Mandarin Chinese-speaking homes if they treated the separate cries as independent within and across infants. However, when they took into account that individual infants have unique and identifiable signatures and statistically nested cries within individuals, the cross-language differences were no longer significant.

In addition, one cannot simply increase size to improve power and assume that you are not, in parallel, also increasing the heterogeneity of your sample. The two factors are not uncoupled. Thus, Kenny and Judd (2019) caution that replication studies with a large sample size may not be representative or definitive unless you explicitly take heterogeneity into account. One approach is to increase N but screen vigorously to ensure homogeneity. However, this is a difficult approach since there are unlikely to be large readily available samples that match, and only match, your core factors of interest.

The alternative is to increase the sample, but also measure and then account for heterogeneity in mechanisms that may influence performance in the task.

As a next step, researchers must then predict how much data the study is likely to yield from the enrolled participants. Most infant researchers have felt the frustration of painstakingly designing a task to address a question of interest, expending time, energy, and resources to train staff and recruit participants, only to have these very participants refuse to, well, participate. They cry. They fuss. They do not calibrate or show no signs of habituating. They are much more interested in the study equipment than any carefully curated set of stimuli. It is frustrating.

And yet, researchers continue in the endeavor, collecting data until they reach an agreed upon sample size (or, the time and money run out). When researchers sit to analyze the data, the unspoken assumption is that the data loss—which can exceed 50%—is missing at random (Enders, 2013). Thus, they treat the data they do have as an unbiased reflection of the population and construct of interest. However, without an explicit check for whether or not the data are missing at random, researchers cannot be sure that this is indeed the case. Of course, we can never be entirely sure, since our assessments are limited to the measures that we took the time to gather and catalog (Little & Rhemtulla, 2013). However, we can at least lower the chances that we are overlooking systematic bias in our sample by incorporating a broader array of measures to help characterize the sample.

6.3 | Creating scientific collaborations

As a science, infant researchers can benefit from shifting our focus to collaborations, both big and small, that can help overcome concerns with individual differences, heterogeneity, and statistical power. These collaborations can grow organically through semi-serendipitous interactions between researchers, or they can be actively built through systematic calls across the field. We have seen successful examples of each in the last few years—three are highlighted here.

First, our current longitudinal study (Pérez-Edgar, LoBue, & Buss, 2017) leverages complementary skills in cognitive development, developmental psychopathology, temperament, and psychophysiology were well suited to studying the early rise of affect-biased attention and links to socioemotional trajectories (Burris et al., 2019). Our research bases, reaching into three distinct locations—State College, PA, Harrisburg, PA, and Newark, NJ—allow us to increase the sample size and, given patterns of ethnic and socioeconomic segregation throughout much of the United States, incorporate a broader range of diversity than typically seen in our communities.

Recently, two larger-scale studies have systematically leveraged strengths and expertise across national and international borders. The first example is the play and learning across a year (PLAY) initiative. Led by three researchers with interwoven expertise (Adolph, Gilmore, & Tamis-LeMonda, 2019) the study will serve as a model system for carrying out developmental science from a "big data" approach. The study will have over 30 laboratories across North America collect 900 hr of video as children engage in naturalistic and semi-structured interactions with a parent. The pooled video (the raw data) will then be distributed to an additional 30 plus laboratories that will code for motor, language, emotional, and behavioral markers of development. The entire corpus will be freely available through Databrary (Gilmore, Adolph, & Millman, 2016) so that the field may leverage the data for their own questions of interest.

Finally, as mentioned above, there is the ManyBabies consortium. Like PLAY, the effort brings together multiple laboratories, this time globally, to collect a pooled set of data. However, each iteration of the study has a shared, fairly singular research focus. Their first multi-laboratory study focused on infant directed speech, which has a robust and long history of research

and fairly nomothetic norms available (ManyBabies Consortium, accepted pending data collection). Reflecting the same approach, ManyBabies 2 (Kampis & Hamlin, 2019) is examining the emergence of theory of mind. This approach can be adjusted to bring individual differences to the center of attention. Of particular interest to the current discussion, ManyBabies 1B (Byers-Heinlein et al., accepted pending data collection) is examining language development in light of variation in infant bilingual exposure. As with PLAY, the focus is on generating a new ethos of open, collaborative science in infant research, creating robust samples that can tackle questions of replication and reproducibility.

7 | CONCLUSION

In her presidential address for the International Congress of Infant Studies (ICIS), Maurer (2015) suggested that infant researchers are often on the hunt for the "superbaby," possessing skills and abilities far beyond the previously assumed limitations inherent in early developmental research. Indeed, the use of looking-time measures, electrophysiology, and clever research designs have revealed that infants notice, process, and interpret a wide array of information across cognitive and socioemotional domains, far beyond their ability to physically act on this information. Patterns of change that emerge as a function of interactions with the environment, and new internal skills, introduce the possibility for individual differences and diverging developmental arcs.

Maurer (2015) noted that incorporating special populations and natural experiments into her research helped identify the depth and breadth of variation in systems of interest. In addition, triangulating across nomothetic and idiographic processes better allowed her research team to identify overlapping and associated critical periods. Examining "eternal laws of human development" requires and integration of both nomothetic and idiographic research (Scarr, 1992, p. 1). This integrative approach allows researchers to examine how each child constructs their own reality from the opportunities afforded by the environment, in the context of individual traits. Noticing and documenting a psychological structure allows researchers to document its presence. By examining variation in the relation between these structures and reactive or regulatory behaviors, researchers can then see the limits of function. In combination, larger-scale studies that incorporate multiple measures and embed experimental methods within observations of naturalistic traits can generate more multi-dimensional views of the emerging infant.

CONFLICTS OF INTEREST

The authors declare no conflicts of interest with regard to the funding source for this study.

ORCID

Koraly Pérez-Edgar https://orcid.org/0000-0003-4051-9563
Alicia Vallorani https://orcid.org/0000-0003-2122-5690
Kristin A. Buss https://orcid.org/0000-0001-5692-6126
Vanessa LoBue https://orcid.org/0000-0002-4386-3549

REFERENCES

Adolph, K. E., Gilmore, R. O., & Tamis-LeMonda, C. (2019). Play & Learning Across a Year (PLAY). Retrieved from play-project.org

Adolph, K. E., & Robinson, S. R. (2015). Motor development. In R. M. Lerner (Ed.), *Handbook of child psychology and developmental science* (7th ed., Vol. 2, pp. 1–45). Hoboken, NJ: John Wiley & Sons.

Allman, M. J., & Mareschal, D. (2016). Possible evolutionary and developmental mechanisms of mental time travel (and implications for autism). *Current Opinion in Behavioral Sciences*, *8*, 220–225. https://doi.org/10.1016/j.cobeha.2016.02.018

Aslin, R. N., & Fiser, J. (2005). Methodological challenges for understanding cognitive development in infants. *Trends in Cognitive Sciences*, *9*, 92–98. https://doi.org/10.1016/j.tics.2005.01.003

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412. https://doi.org/10.1016/j.jml.2007.12.005

Barnett, W. S. (2011). Effectiveness of early educational intervention. *Science*, *333*(6045), 975–978.

Behrmann, M., Avidan, G., Leonard, G. L., Kimchi, R., Luna, B., Humphreys, K., & Minshew, N. (2006). Configural processing in autism and its relationship to face processing. *Neuropsychologia*, *44*(1), 110–129. https://doi.org/10.1016/j.neuropsychologia.2005.04.002

Bergmann, C., Tsuji, S., Piccinini, P. E., Lewis, M. L., Braginsky, M., Frank, M. C., & Cristia, A. (2018). Promoting replicability in developmental research through meta-analyses: Insights from language acquisition research. *Child Development*, *89*(6), 1996–2009. https://doi.org/10.1111/cdev.13079

Bornstein, M. H., Arterberry, M. E., & Mash, C. (2011). Perceptual development. In M. H. Bornstein & M. E. Lamb (Eds.), *Developmental psychology: An advanced textbook* (6th ed., pp. 311–360). New York, NY: Psychology Press.

Bremner, J. G., Johnson, S. P., Slater, A., Mason, U., Foster, K., Cheshire, A., & Spring, J. (2005). Conditions for young infants' perception of object trajectories. *Child Development*, *76*(5), 1029–1043. https://doi.org/10.1111/j.1467-8624.2005.00895.x

Bremner, J. G., Slater, A. M., Mason, U. C., Spring, J., & Johnson, S. P. (2017). Limits of object persistence: Young infants perceive continuity of vertical and horizontal trajectories, but not 45-Degree oblique trajectories. *Infancy*, *22*(3), 303–322. https://doi.org/10.1111/infa.12170

Burris, J. L., Oleas, D., Reider, L., Buss, K. A., Pérez-Edgar, K., & LoBue, V. (2019). Biased attention to threat: Answering old questions with young infants. *Current Directions in Psychological Science*, *28*, 534–539. https://doi.org/10.1177/0963721419861415

Buss, K. A. (2011). Which fearful toddlers should we worry about? Context, fear regulation, and anxiety risk. *Developmental Psychology*, *47*, 804–819. https://doi.org/10.1037/a0023227

Buss, K. A., Davis, E. L., Ram, N., & Coccia, M. (2018). Dysregulated fear, social inhibition, and respiratory sinus arrhythmia: A replication and extension. *Child Development*, *89*(3), e214–e228.

Buss, K. A., & Qu, J. (2018). Psychobiological processes in the development of behavioral inhibition. In K. Pérez-Edgar & N. A. Fox (Eds.), *Behavioral inhibition* (pp. 91–111). Cham, Switzerland: Springer.

Byers-Heinlein, K., Bergmann, C., Black, A., Carbajal, J. M., Fennell, C. T., Frank, M. C., … Tsui, A. S. M. (accepted pending data collection). A multi-lab study of bilingual infants: Exploring the preference for infant-directed speech. *Advances in Methods and Practices in Psychological Science*.

Carey, S., & Spelke, E. (1994). Domain-specific knowledge and conceptual change. *Mapping the Mind: Domain Specificity in Cognition and Culture*, *169*, 200.

Carroll, J. B. (1978). How shall we study individual differences in cognitive abilities?-Methodological and theoretical perspectives. *Intelligence*, *2*(2), 87–115. https://doi.org/10.1016/0160-2896(78)90002-8

Chaplin, T. M., & Aldao, A. (2013). Gender differences in emotion expression in children: A meta-analytic review. *Psychological Bulletin*, *139*, 735–765. https://doi.org/10.1037/a0030737

Chaplin, T. M., Cole, P. M., & Zahn-Waxler, C. (2005). Parental socialization of emotion expression: Gender differences and relations to child adjustment. *Emotion*, *5*, 80. https://doi.org/10.1037/1528-3542.5.1.80

Dawson, G., Webb, S. J., & McPartland, J. (2005). Understanding the nature of face processing impairment in autism: Insights from behavioral and electrophysiological studies. *Developmental Neuropsychology*, *27*(3), 403–424. https://doi.org/10.1207/s15326942dn2703_6

Deater–Deckard, K., Dodge, K. A., Bates, J. E., & Pettit, G. S. (1998). Multiple risk factors in the development of externalizing behavior problems: Group and individual differences. *Development and Psychopathology*, *10*(3), 469–493.

DeBolt, M. C., Rhemtulla, M., & Oakes, L. M. (2020, under review). Robust data and power in infant research: A case study of the effect of number of infants and number of trials in visual preference procedures. *Infancy*, 1–27. https://doi.org/10.1111/infa.12337

Diamond, A., Cruttenden, L., & Neiderman, D. (1994). AB with multiple wells: 1. Why are multiple wells sometimes easier than two wells? 2. Memory or memory+inhibition? *Developmental Psychology*, *30*, 192–205.

Duffy, R. D., Jadidian, A., Webster, G. D., & Sandell, K. J. (2011). The research productivity of academic psychologists: Assessment, trends, and best practice recommendations. *Scientometrics*, *89*(1), 207–227. https://doi.org/10.1007/s11192-011-0452-4

Eason, A. E., Hamlin, J. K., & Sommerville, J. A. (2017). A survey of common practices in infancy research: Description of policies, consistency across and within labs, and suggestions for improvements. *Infancy*, *22*(4), 470–491. https://doi.org/10.1111/infa.12183

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*, 175–191. https://doi.org/10.3758/BF03193146

Fox, N. A., & Pine, D. S. (2012). Temperament and the emergence of anxiety disorders. *Journal of the American Academy of Child and Adolescent Psychiatry*, *51*, 125–128. https://doi.org/10.1037//0021-843X.103.1.18

Fox, N. A., Snidman, N., Haas, S. A., Degnan, K. A., & Kagan, J. (2015). The relations between reactivity at 4 months and behavioral inhibition in the second year: Replication across three independent samples. *Infancy*, *20*(1), 98–114.

Frank, M. C., Bergelson, E., Bergmann, C., Cristia, A., Floccia, C., Gervain, J., … Yurovsky, D. (2017). A collaborative approach to infant research: Promoting reproducibility, best practices, and theory-building. *Infancy*, *22*(4), 421–435. https://doi.org/10.1111/infa.12182

Friedlmeier, W., & Trommsdorff, G. (1999). Emotion regulation in early childhood: A cross-cultural comparison between German and Japanese toddlers. *Journal of Cross-Cultural Psychology*, *30*(6), 684–711. https://doi.org/10.1177/0022022199030006002

Gilmore, R. O., Adolph, K. E., & Millman, D. S. (2016). *Curating identifiable data for sharing: The databrary project*. Paper presented at the 2016 New York Scientific Data Summit (NYSDS).

Giżewska, M. (2015). Phenylketonuria: Phenylalanine neurotoxicity. In L. E. Bernstein, F. Rohr, J. R. Helm (Eds.), *Nutrition management of inherited metabolic diseases* (pp. 89–99). Cham, Switzerland: Springer.

Goodhew, S. C., & Edwards, M. (2019). Translating experimental paradigms into individual-differences research: Contributions, challenges, and practical recommendations. *Consciousness and Cognition*, *69*, 14–25. https://doi.org/10.1016/j.concog.2019.01.008

Guldberg, P., Rey, F., Zschocke, J., Romano, V., François, B., Michiels, L., … Güttler, F. (1998). A European multicenter study of phenylalanine hydroxylase deficiency: Classification of 105 mutations and a general system for genotype-based prediction of metabolic phenotype. *The American Journal of Human Genetics*, *63*(1), 71–79. https://doi.org/10.1086/301920

Gustafson, G. E., Sanborn, S. M., Lin, H. C., & Green, J. A. (2017). Newborns' cries are unique to individuals (but not to language environment). *Infancy*, *22*(6), 736–747. https://doi.org/10.1111/infa.12192

Hunter, M. A., & Ames, E. W. (1988). A multifactor model of infant preferences for novel and familiar stimuli. In C. Rovee-Collier & L. P. Lipsitt (Eds.), *Advances in Infancy Research*, (Vol. *5*, pp. 69–95). Norwood, NJ: Ablex Publishing.

Iverson, J. M. (2010). Developing language in a developing body: The relationship between motor development and language development. *Journal of Child Language*, *37*(2), 229–261. https://doi.org/10.1017/S0305000909990432

Izard, C. E., Hembree, E. A., & Huebner, R. R. (1987). Infants' emotion expressions to acute pain: Developmental change and stability of individual differences. *Developmental Psychology*, *23*(1), 105. https://doi.org/10.1037/0012-1649.23.1.105

Johnson, M. H. (2010). Developmental neuroscience, psychophysiology, and genetics. In M. H. Bornstein & M. E. Lamb (Eds.), *Developmental science: An advanced textbook* (pp. 209–248). Mahwah, NJ: Lawrence Erlbaum Associates.

Kagan, J. (2018a). The bases for preservation of emotional biases. In A. S. Fox, R. C. Lapate, A. J. Shackman, & R. J. Davidson (Eds.), *The nature of emotion: Fundamental questions* (2nd ed., pp. 64–67). New York, NY: Oxford University Press.

Kagan, J. (2018b). The history and theory of behavioral inhibition. In K. Pérez-Edgar & N. A. Fox (Eds.), *Behavioral inhibition* (pp. 1–15). Cham, Switzerland: Springer.

Kampis, D., & Hamlin, K. (2019). *ManyBabies 2: A multi-lab study on infant theory of mind*. Paper presented at the Society for Research in Child Development, Baltimore, MD.

Karasik, L. B., Adolph, K. E., Tamis-LeMonda, C. S., & Bornstein, M. H. (2010). WEIRD walking: Cross-cultural research on motor development. *Behavioral and Brain Sciences*, *33*(2–3), 95–96. https://doi.org/10.1017/S0140525X10000117

Kenny, D. A., & Judd, C. M. (2019). The unappreciated heterogeneity of effect sizes: Implications for power, precision, planning of research, and replication. *Psychological Methods*, *24*(5), 578–589. https://doi.org/10.1037/met0000209

Kiel, E. J., & Kalomiris, A. E. (2015). Current themes in understanding children's emotion regulation as developing from within the parent–child relationship. *Current Opinion in Psychology*, *3*, 11–16. https://doi.org/10.1016/j.copsyc.2015.01.006

Kuchirko, Y., Tafuro, L., & Tamis LeMonda, C. S. (2018). Becoming a communicative partner: Infant contingent responsiveness to maternal language and gestures. *Infancy*, *23*(4), 558–576. https://doi.org/10.1111/infa.12222

Lamb, M. E. (2015). Processes underlying social, emotional, and personality development: A preliminary survey of the terrain. In R. M. Lerner (Ed.), *Handbook of child psychology and developmental science* (pp. 1–10). Hoboken, NJ: Wiley.

Leonard, H. C., & Hill, E. L. (2014). The impact of motor development on typical and atypical social cognition and language: A systematic review. *Child and Adolescent Mental Health*, *19*(3), 163–170.

Lewis, M. D. (2000). The promise of dynamic systems approaches for an integrated account of human development. *Child Development*, *71*(1), 36–43. https://doi.org/10.1111/1467-8624.00116

Little, T. D., & Rhemtulla, M. (2013). Planned missing data designs for developmental researchers. *Child Development Perspectives*, *7*(4), 199–204. https://doi.org/10.1111/cdep.12043

LoBue, V., Buss, K. A., Taber-Thomas, B. C., & Pérez-Edgar, K. (2017). Developmental differences in infants' attention to social and non-social threats. *Infancy*, *22*, 403–415. https://doi.org/10.1111/infa.12167

LoBue, V., Kim, E., & Delgado, M. R. (2019). Fear in development. In V. LoBue, K. Pérez-Edgar, & K. A. Buss (Eds.), *Handbook of emotional development* (pp. 257–282). Cham, Switzerland: Springer.

LoBue, V., Reider, L., Kim, E., Burris, J. L., Oleas, D. S., Buss, K. A., … Field, A. (2020). The importance of using multiple outcome measures in infant research. *Infancy*, 1–18. https://doi.org/10.1111/infa.12339

MacNeill, L., Ram, N., Bell, M. A., Fox, N. A., & Pérez-Edgar, K. (2018). Trajectories of infants' biobehavioral development: Timing and rate of A-not-B performance gains and EEG maturation. *Child Development*, *89*, 711–724. https://doi.org/10.1111/cdev.13022

ManyBabies Consortium. (accepted pending data collection). Quantifying sources of variability in infancy research using the infant-directed speech preference. *Advances in Methods and Practices in Psychological Science*.

Mareschal, D. (2000). Infant object knowledge: Current trends and controversies. *Trends in Cognitive Science*, *4*, 408–416.

Mareschal, D., & French, R. (2000). Mechanisms of categorization in infancy. *Infancy*, *1*(1), 59–76. https://doi.org/10.1207/S15327078IN0101_06

Martins, C., & Gaffan, E. A. (2000). Effects of early maternal depression on patterns of infant–mother attachment: A meta-analytic investigation. *The Journal of Child Psychology and Psychiatry and Allied Disciplines*, *41*(6), 737–746. https://doi.org/10.1111/1469-7610.00661

Maurer, D. (2015). What atypical adults can teach us about development. *Infancy*, *20*(6), 587–600. https://doi.org/10.1111/infa.12106

McCall, R. B. (1977). Challenges to a science of developmental psychology. *Child Development*, 333–344. https://doi.org/10.2307/1128626

McCartney, K., & Rosenthal, R. (2000). Effect size, practical importance, and social policy for children. *Child Development*, *71*(1), 173–180. https://doi.org/10.1111/1467-8624.00131

Mills-Smith, L., Spangler, D. P., Panneton, R., & Fritz, M. S. (2015). A missed opportunity for clarity: Problems in the reporting of effect size estimates in infant developmental science. *Infancy*, *20*(4), 416–432. https://doi.org/10.1111/infa.12078

Muthén, L., & Muthén, B. (2002). How to use a Monte Carlo study to decide on sample size and determine power. *Structural Equation Modeling*, *4*, 599–620. https://doi.org/10.1207/S15328007SEM0904_8

Nelson, C. A. (2001). The development and neural bases of face recognition. *Infant and Child Development*, *10*(1–2), 3–18. https://doi.org/10.1002/icd.239

Noble, K. G., Houston, S. M., Kan, E., & Sowell, E. R. (2012). Neural correlates of socioeconomic status in the developing human brain. *Developmental Science*, *15*(4), 516–527. https://doi.org/10.1111/j.1467-7687.2012.01147.x

Nomi, J. S., & Uddin, L. Q. (2015). Face processing in autism spectrum disorders: From brain regions to brain networks. *Neuropsychologia*, *71*, 201–216. https://doi.org/10.1016/j.neuropsychologia.2015.03.029

Oakes, L. M. (2017). Sample size, statistical power, and false conclusions in infant looking-time research. *Infancy*, *22*, 436–469. https://doi.org/10.1111/infa.12186

Peltola, M. J., Forssman, L., Puura, K., IJzendoorn, M. H., & Leppänen, J. M. (2015). Attention to faces expressing negative emotion at 7 months predicts attachment security at 14 months. *Child Development*, *86*(5), 1321–1332.

Peltola, M. J., Leppanen, J. M., Maki, S., & Hietanen, J. K. (2009). Emergence of enhanced attention to fearful faces between 5 and 7 months of age. *Social Cognitive and Affective Neuroscience*, *4*, 134–142. https://doi.org/10.1093/scan/nsn046

Peltola, M. J., Leppanen, J. M., Palokangas, T., & Hietanen, J. K. (2008). Fearful faces modulate looking duration and attention disengagement in 7-month-old infants. *Developmental Science*, *11*, 60–68. https://doi.org/10.1111/j.1467-7687.2007.00659.x

Peltola, M. J., Leppanen, J. M., Vogel-Farley, V. K., Hietanen, J. K., & Nelson, C. A. (2009). Fearful faces but not fearful eyes alone delay attention disengagement in 7-month-old infants. *Emotion*, *9*, 560–565. https://doi.org/10.1037/a0015806

Peltola, M. J., Yrttiaho, S., & Leppänen, J. M. (2018). Infants' attention bias to faces as an early marker of social development. *Developmental Science*, *21*(6), e12687. https://doi.org/10.1111/desc.12687

Pérez-Edgar, K., & Hastings, P. D. (2018). Emotion development from an experimental and individual differences lens. In J. T. Wixted & S. Ghetti (Eds.), *Stevens' handbook of experimental psychology and cognitive neuroscience* (4th ed., Vol. *4*, pp. 289–321). New York, NY: Wiley.

Pérez-Edgar, K., LoBue, V., & Buss, K. A. (2017). LANTS: Longitudinal Attention and Temperament Study. from Databrary, https://nyu.databrary.org/volume/485

Pérez-Edgar, K., Morales, S., LoBue, V., Taber-Thomas, B. C., Allen, E. K., Brown, K. M., & Buss, K. A. (2017). The impact of negative affect on attention patterns to threat across the first two years of life. *Developmental Psychology*, *53*, 2219–2232.

Rademaker, R. L., Park, Y. E., Sack, A. T., & Tong, F. (2018). Evidence of gradual loss of precision for simple features and complex objects in visual working memory. *Journal of Experimental Psychology: Human Perception and Performance*, *44*(6), 925.

Romberg, A. R., & Saffran, J. R. (2010). Statistical learning and language acquisition. *Wiley Interdisciplinary Reviews: Cognitive Science*, *1*(6), 906–914. https://doi.org/10.1002/wcs.78

Rubio-Fernández, P. (2019). Publication standards in infancy research: Three ways to make violation-of-expectation studies more reliable. *Infant Behavior and Development*, *54*, 177–188. https://doi.org/10.1016/j.infbeh.2018.09.009

Scarr, S. (1992). Developmental theories for the 1990s: Development and individual differences. *Child Development*, *63*, 1–19. https://doi.org/10.2307/1130897

Schott, E., Rhemtulla, M., & Byers-Heinlein, K. (2019). Should I test more babies? Solutions for transparent data peeking. *Infant Behavior and Development*, *54*, 166–176. https://doi.org/10.1016/j.infbeh.2018.09.010

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, *22*(11), 1359–1366. https://doi.org/10.1177/0956797611417632

Van Geert, P. (1998). A dynamic systems model of basic developmental mechanisms: Piaget, Vygotsky, and beyond. *Psychological Review*, *105*(4), 634. https://doi.org/10.1037/0033-295X.105.4.634-677

Wellman, H. M., Cross, D., & Bartsch, K. (1987). Infant search and object permanence: A meta-analysis of the A-not-B error. *Monographs of the Society for Research in Child Development*, *51*(3), 1–67.

Wohlwill, J. F. (1973). *The study of behavioral development*. New York, NY: Academic Press.

Xu, F. (2003). Numerosity discrimination in infants: Evidence for two systems of representations. *Cognition*, *89*(1), B15–B25. https://doi.org/10.1016/S0010-0277(03)00050-7

How to cite this article: Pérez-Edgar K, Vallorani A, Buss KA, LoBue V. Individual differences in infancy research: Letting the baby stand out from the crowd. *Infancy*. 2020;25:438–457. https://doi.org/10.1111/infa.12338