The Global Temperament Project: Parent-Reported Temperament in Infants, Toddlers and Children from 59 Nations

Samuel P. Putnam

Bowdoin College

Ela Sehic, Brian F. French, Maria A. Gartstein

Washington State University

Benjamin Lira Luttges

University of Pennsylvania

486 Additional Partners in the Global Temperament Project

Correspondence concerning this article should be addressed to Samuel Putnam, Department of Psychology, Bowdoin College, Brunswick, ME 04011. Phone: 207-725-3152. Email: sputnam@bowdoin.edu. Samuel Putnam carried out roles of conceptualization, analysis, writing, administration and visualization. Ela Sehic and Brian French contributed to data analysis. Maria Gartstein contributed to conceptualization and writing. Benjamin Lira Luttges contributed to data visualization. All other coauthors contributed to investigation or data curation and are listed alphabetically in Appendix 1. Due to intellectual property agreements with participating sites, raw data are not publicly available, although summary files will be made available upon request to the corresponding author.

Abstract

Data from 83,423 parent reports of temperament (Surgency, Negative Affectivity, and Regulatory Capacity) in infants, toddlers and children from 341 samples gathered in 59 countries were used to investigate the relations among culture, gender and temperament. Between-nation differences in temperament were larger than those obtained in similar studies of adult personality, and most pronounced for Negative Affectivity. Nation-level patterns of Negative Affectivity were consistent across infancy, toddlerhood and childhood; and patterns of Regulatory Capacity were consistent between infancy and toddlerhood. Nations that previously reported high Extraversion, high Conscientiousness, and low Neuroticism in adults were found to demonstrate high Surgency in infants and children; and countries reporting low adult Openness and high adult Neuroticism reported high temperamental Negative Affectivity. Negative Affectivity was high in Southern Asia, Western Asia and South America; and low in Northern and Western Europe. Countries in which children were rated as high in Negative Affectivity had cultural orientations reflecting Collectivism, high Power Distance, and Short-Term Orientation. Surgency was high in Southeastern and Southern Asia and Southern Europe, and low in Eastern Asian countries characterized by philosophies of Long-Term Orientation. Low personal income was associated with high Negative Affectivity. Gender differences in temperament were largely consistent in direction with prior studies, revealing higher Regulatory Capacity in females than males and higher Surgency in males than females; with these differences becoming more pronounced at later ages.

Keywords: Temperament, culture, gender, personality, infants, children

Public Significance Statement: This study is the largest existing effort (59 nations and 83,423 parent reports) to document and understand between-nation differences in the social and emotional behavior of infants and young children. The results suggest that children in Collectivist nations of South America and Southern Asia expressed more negative emotions than those from Northern and Western Europe. Gender differences were relatively consistent across nations and grew stronger with increasing age.

The Global Temperament Project: Parent-Reported Temperament in Infants, Toddlers and Children from 59 Nations

When considering the interminable question of how nature and nurture shape individual differences in humans, temperament and culture may be viewed as representing opposite poles. The concept of temperament emphasizes forces of nature, with characteristic tendencies to approach and react to stimuli viewed as being influenced by biological mechanisms inherent to the individual (e.g., Allport, 1961). In contrast, culture emphasizes the role of nurture, with behavior continually shaped by repeated communications with members of society (e.g., Vygotsky, 1978), as overarching "macrosystem" values and beliefs shared by social groups are communicated through "microsystem" interactions between children and their social companions (Bronfenbrenner, 1979). Given their conceptual distance in classical perspectives, one might expect temperament and culture to be largely unrelated to one another, each providing independent influence on the developing person.

Contemporary perspectives on development, however, recognize the falseness of the nature-nurture dichotomy, viewing development as dependent on dynamic, bidirectional processes taking place between multiple levels of the person and their environment (e.g., Burman, 2019; Gottlieb, 1991). The definitions of temperament have evolved accordingly: whereas early theories listed heritability as a defining criterion (e.g., Buss & Plomin, 1975), newer approaches define temperament traits as "early emerging basic dispositions in the domains of activity, affectivity, attention and self-regulation, and these dispositions are the product of complex interactions among genetic, biological *and environmental factors* across time" (Shiner et al., 2012, p. 437; [emphasis added). This definition suggests that repeated cultural differences in the treatment of children may influence their behavior in predictable ways. Conversely, it is recognized that demands of culture may influence the genetic composition of populations through natural selection and that the emotional and social proclivities of groups will shape cultural norms (Boyd & Richerson, 1985; Chiao & Blizinsky, 2010; Putnam & Gartstein, 2018; Rentfrow et al., 2008). Consistent with increasing appreciation of the reciprocal effects of environments, biological factors, and behavioral systems upon one another, a growing body of research has documented empirical relations between culture and temperament. To date, these studies have been limited in scope, involving small samples of children at a given developmental period from only a few countries. The current investigation expands upon these efforts by providing the most comprehensive examination to date of relations between culture and individual differences in reactivity and regulation, exploring patterns of temperament around the globe through analyses of 83,423 parent reports of behavior in infants, toddlers, and children from 59 countries.

Temperament Measurement and Structure

The measures used in the current study emerged from the psychobiological model of temperament proposed by Rothbart (e.g., Rothbart, 2011; Rothbart & Derryberry, 1981), who defined temperament as early appearing, relatively stable, constitutionally based individual differences in reactivity and regulation, influenced over time by genetics, experience and maturation. Reactivity applies to the threshold, magnitude and latency of responses made by emotional, sensory and motor systems to relevant stimuli, while regulation refers to attentional and inhibitory processes that modulate reactions. Referring to this definition and drawing upon conceptual and psychometric evaluations of early questionnaire measures (e.g., Bates et al., 1979; Derryberry & Rothbart, 1988; Martin et al., 1988; Rothbart, 1981; Thomas & Chess, 1977) and relevant concepts from animal and adult personality, Rothbart and colleagues created the

Children's Behavior Questionnaire (CBQ; Rothbart et al., 2001). The CBQ was more highly differentiated than its predecessors, comprising internally consistent scales for 15 distinct aspects of temperament to be measured in children between the ages of three and eight years. The rational approach to scale development used to develop the CBQ was subsequently followed to construct analogous fine-grained measures for use with infants (Infant Behavior Questionnaire-Revised; IBQ-R; Gartstein & Rothbart, 2003), toddlers (Early Childhood Behavior Questionnaire; ECBQ; Putnam et al., 2006), older children (Simonds, 2006), adolescents (Ellis & Rothbart, 2001) and adults (Evans & Rothbart, 2007). Three of these measures – the IBQ-R, ECBQ and CBQ – are used in the current study.

The use of parent reports as measures of child temperament has been widely critiqued (e.g., Kagan, 1994). Parents may be influenced by social desirability, exaggerating what they perceive to be socially desirable qualities and underestimating undesired conduct. The validity of parents' ratings may be further compromised by the influence of personality characteristics, transitory states, differing interpretations of questions and insufficient comparison groups (Rothbart & Bates, 2006). The authors of the psychobiological battery of questionnaires have attempted to minimize these concerns by writing items regarding the frequency of concrete behaviors in commonly occurring situations, but the potential for bias – in culturally specific directions – remains a concern, a point to which we return in the discussion. Despite their shortcomings, parent reports are uniquely comprehensive as a source of information, particularly with respect to young children. Parents' unique position allows them to view their child in countless contexts, including those that are logistically or ethically impossible in lab settings. Caregivers also observe their offspring over multiple trials, enabling them to separate rare responses from more typical behaviors for their children.

The wide range of behaviors considered in parent report questionnaires of the psychobiological approach to temperament also allows for explorations of higher-order organization of traits. Although the items and scales making up these age-based measures are considerably different, factor analyses of scale scores have revealed a relatively similar factor structure (e.g., Putnam et al., 2001). At all ages, a dimension labeled Negative Emotionality, involving tendencies to experience and display fear, anger, sadness and physical discomfort, has been related conceptually and empirically to Big Five and Five Factor Model Neuroticism (e.g., Evans & Rothbart, 2007; Shiner & Caspi, 2003). Surgency emerges as a second factor, primarily manifested through smiling, laughing, activity, appreciation of high intensity stimulation and approaching novel stimuli; it corresponds to Big Five Extraversion. A third factor, referred to as Regulatory Capacity in infancy and later as Effortful Control includes attentional abilities, behavioral control, and enjoyment of calm activities, with associations to Conscientiousness. For efficiency, in the current manuscript, we use the term Regulatory Capacity to refer to this trait, regardless of the age at which it was measured. Very Short Forms of the IBQ-R, ECBQ and CBQ have been developed, each containing three scales intended to measure these broad factors (Putnam et al., 2014; Putnam et al., 2010; Putnam & Rothbart, 2006). Although item-level analyses have revealed alternate structures for these instruments (Kotelnikova et al., 2016; Peterson et al., 2017), the replicability of the three-factor framework of Surgency, Negative Affectivity and Regulatory Capacity has been supported by several cross-cultural and psychometric investigations (e.g., Golmohammadi et al., 2020; Gartstein et al., 2005; Montirosso et al., 2011; Sleddens et al., 2011).

Temperament and Culture

Two recent frameworks have been particularly influential in characterizing the processes through which culture is translated to patterns of variation in temperament. The notion of the developmental niche (Harkness & Super, 1994; Super & Harkness, 1986) views culture as shaping children's environments through its influence on the physical and social settings in which they develop and the customary ways that they are treated, which are governed by parental ethnotheories (i.e., beliefs regarding how children are to be treated and the qualities they expect children to develop early in life). Complementing the concept of the developmental niche, the contextual-developmental perspective proposed by Chen (e.g., 2018; Chen & French, 2008; see also Kohnstamm, 1989) emphasizes the evaluation and response processes enacted by adults and peers. In this model, members of the developing child's world respond to behaviors according to values and expectations inherent to their cultural norms, expressing approval of sanctioned conduct and rejection of inappropriate actions. To the degree that the evaluated behavioral tendencies are amenable to change, children alter behaviors associated with negative or positive social evaluations. For instance, taking social initiative is viewed as a valuable goal of socialization in Western societies that emphasize the self, but undesirable in Eastern societies that are more group oriented (Chen, Liu & Bian, 2022). Views regarding self-control also differ between individualist nations, in which it is promoted as a vehicle for personal achievement and independence, and collectivist cultures, in which restraint in seeking one's own desires is encouraged as a mechanism to maintain group harmony (Lee et al., 2013). Indeed, shyness (representing low social initiative) has been associated with parental warmth and positive familial relationships in China, South Korea, and Thailand, but disappointment and concern among parents in Canada, the United States and Australia (e.g., Chen et al., 1998; Kim et al., 2008). Similarly, parents in China have stronger expectations for self-control in their children than their counterparts in North America (Chen et al., 2003; Lee et al., 2013). Consistent with these parental expectations, both parent-report and observational measures have revealed higher levels of fearfulness and effortful restraint among children from East Asian societies than those from Western cultures (e.g., Chen et al., 1998; Krassner et al., 2017; Rubin et al., 2006). A more limited body of research similarly suggests higher levels of fearfulness and self-control among children from Hispanic/Latino cultures, in comparison to those raised in the US, also revealing considerable variability among children from different Latin American regions (Galindo & Fuller, 2010; Gudino & Lao, 2010; Polo & Lopez, 2009).

A limitation of most studies considering culture in relation to temperament concerns geographic scope. With some exceptions (e.g., Super et al., 2008; Super et al., 2020), comparisons involving infants and children from areas outside East Asia and the Americas are rare. In addition to omitting knowledge from a rich pool of global variability, most studies in this tradition involve only two countries, but are sometimes generalized with respect to an entire region. The differences in self-control between children of Cuban/South American and those of Mexican and Puerto Rican heritage observed by Galindo and Fuller (2010) reveal the limits of considering one country to be representative of a larger region. Relatedly, most studies base their country-level estimations of temperament on single samples from limited geographic and socio-political areas but interpret these estimates as generalizing to their nation of origin. It is known that patterns of personality characteristics vary meaningfully across different parts of the United States (Rentfrow, Gosling & Potter, 2008), suggesting that temperament might similarly differ between regions within a country, and demonstrating the importance of using multiple samples per country when possible. Likewise, most studies involve children from a restricted age range (c.f., Gaias et al., 2012) obscuring potential differences in patterns at other ages.

Two recent cross-cultural comparison studies couched in the psychobiological tradition have addressed some limitations of prior studies. The first (Putnam & Gartstein, 2017) involved a meta-analysis of 17 prior studies of nation-level differences in temperament of infants, children and adults from 18 countries. From the scores presented in the earlier reports, aggregate scores representing Surgency, Negative Affectivity and Regulatory Capacity were derived for each of the countries. These aggregate scores were then analyzed in relation to analogous values from cross-cultural studies of adult personality, dimensions of cultural orientation, and allelic frequency estimates. The second, referred to as the Joint Effort Toddler Temperament Consortium (JETTC; Gartstein & Putnam, 2018), examined ECBQ data collected from 865 families in 14 sites and 12 countries, relating between- and within-culture differences in temperament to patterns of parental philosophy (i.e., socialization goals and parental ethnotheories; Keller et al., 2006), aspects of daily activities, parental responses to temperament displays and children's behavior problems (Achenbach & Rescorla, 2000).

The Present Study

Although the JETTC and meta-analytic investigations expanded the scope of cross-cultural understanding of temperament, they continued to demonstrate several of the limitations described above. The range of countries in both studies was largely limited to large and populous nations. In both, Western Europe was relatively well-represented, but few or no countries from Latin America, Eastern Europe, or Southeast Asia were included. Of the 18 countries in the meta-analysis, scores for eight came from a single sample in that country, and scores for seven came from only two; and in the JETTC, all countries but two were represented by data collected in a single community. Some samples from studies included in the meta-analyses contained several hundred participants, but the majority contained fewer than 100; and the JETTC sites recruited around 50 families each. The Global Temperament Project (GTP) was organized to improve upon these past efforts. The current study uses data from 83,423 parent reports of temperament in infants, toddlers and children from 341 samples gathered in 59 countries to investigate *six questions* regarding relations between culture and temperament.

The first question simply concerns the magnitude of between-nation differences in temperament. The restricted range of previous studies has limited the validity of previous estimates of the proportion of variance associated with between-nation, relative to within-nation, differences between individuals. The data compiled for the GTP allow for a comparison between effect sizes for nation and those evident for two other central sources of variation: age and gender. Variants on this question involve comparisons across age and construct. A larger effect of nation at later ages than earlier ages would provide support for compounding influence of cultural context across the first several years of life. Differences in the size of effects among dimensions of Surgency, Negative Affectivity and Regulatory Capacity can reveal the degree to which some early-appearing aspects of reactivity and regulation manifest themselves in ways consistent with cultural demands than others.

In addition to providing a useful comparison point for relations between temperament and culture, child gender forms the basis of *the second question to be addressed with these data: the relative universality of gender differences* in infancy and early childhood. To date, the most thorough examination of temperament in relation to gender is a meta-analysis conducted by Else-Quest et al. (2006).[1] This quantitative summary of 205 studies of children ages 3 months to 13 years revealed moderate gender differences reflecting higher Regulatory Capacity in girls and higher Surgency in boys, with only small gender differences for Negative Affectivity in the narrow dimensions of fear (higher in girls) and "difficulty" (higher in boys). We anticipate similar main effects of child gender in the current study.

Predictions regarding nation as a moderator of gender on temperament are more elusive. Although the samples utilized by Else-Quest et al. (2006) represented a range of nationalities, country of origin was not explored as a potential moderator of gender differences. Interactions between gender and nation-level differences have been reported in individual studies of temperament, but the combined results of these studies demonstrate no discernable pattern. For instance, Montirosso et al. (2011) found Italian male, but not female, infants to be rated higher in cuddliness than their US counterparts; while Cozzi et al. (2013) found Italian male toddlers to be higher in soothability than US male toddlers.

Similarities described above between the three psychobiological temperament dimensions and Big Five personality traits suggest that cross-cultural studies of personality may inform predictions regarding gender and culture. Early studies in this vein suggested that patterns of gender in relation to personality were consistent across nations (see Feingold, 1994), and some recent large studies have also suggested considerable cross-cultural consistency in the direction of gender effects for adult personality (Kajonius & Johnson, 2018; McCrae & Terracciano, 2005) and in the timing of the emergence of gender differences across adolescence (De Bolle et al., 2015). The size of such differences, however, varies considerably. Kaiser (2019) replicated and extended previous studies (e.g., Costa et al., 2001; Schmitt et al., 2008) linking gender differences to attributes of culture, generating support for the “Gender Equality Paradox”, in which personality differences between men and women tend to be greater in more gender-egalitarian, individualist, and developed countries. The possibility that these social forces act in a similar manner upon gender differences in infancy and young children is intriguing, with potential implications for the role of society in shaping gendered behavior.

A third question, which is allowed by the inclusion of data across multiple developmental periods and correspondingly divergent measures in the GTP, concerns the *relative consistency of nation-level differences in temperament across infancy, toddlerhood and childhood.* The psychobiological model emphasizes the importance of development in molding early appearing individual differences (e.g., Rothbart, 2011). With increasing age, biological maturation is complemented by changing expectations regarding the expression of emotions and control of one's attention and impulses. The age at which certain behaviors are expected or allowed differs across societies, such that high levels of a given behavioral tendency in a culture during infancy may not predict similar levels of analogous behaviors at older ages. In addition, although longitudinal studies have documented relative continuity of Surgency, Negative Affectivity and Regulatory Capacity from infancy through childhood, this stability is relatively modest. For instance, Putnam et al. (2008) reported longitudinal correlations for the three factors that ranged from .34 to .36 from infancy to toddlerhood, and .49 to .59 from toddlerhood to early childhood, but found no correspondence between Regulatory Capacity measured with the IBQ-R and CBQ, and correlations of only .25 and .36 for Surgency and Negative Affectivity across this developmental span. Furthermore, connections between behavioral tendencies are not straightforward, but instead exhibit heterotypic continuity and cascades. Of particular note, Putnam et al. (2008) found ECBQ Regulatory Capacity to be predicted by high levels of IBQ Surgency.

A few existing studies have indicated consistency of nation-level differences in temperament across the lifespan. Gaias et al. (2012) found higher fearfulness in US than Finnish individuals during infancy, childhood and adulthood; Montirosso et al. (2011) and Cozzi et al. (2013) reported higher Cuddliness (a component of Regulatory Capacity) and lower High Intensity Pleasure (a component of Surgency) in Italian than US individuals at both infancy and toddlerhood; and Slobodskaya et al. (2013) had consistent results from infants and toddlers from Russia, Japan and the US, with US youngsters scoring higher on Surgency than Russian and Japanese kids, and Japanese youth viewed as lower on aspects of Regulatory Capacity at both ages. In the current effort, the GTP will expand upon these limited comparisons of two or three nations to more formally explore consistency in nation-level differences across early development.

The consistency of relations between culture and individual differences across wider spans in terms of both human development and methodology forms the basis of *our fourth question: Do nation-level differences in temperament resemble those found for adult personality?* Although no large-scale investigations of culture and temperament precede the current report, several multi-nation studies involving adult and adolescent personality have been published. A landmark report on self-reported five-factor scores in 26 countries by McCrae (2001) was soon supplemented to include 36 countries (McCrae, 2002). Allik et al. (2017) recently combined these 36 aggregate scores with data from other published and unpublished studies to develop aggregate personality scores for 62 countries and 76 samples. Country-level estimates of observer reports have also been created. McCrae et al. (2005) asked college students in 51 cultures to complete ratings of a friend from their country. These other-report values were roughly consistent with the self-report scores generated by McCrae (2002), as scores for Neuroticism and Extraversion converged modestly with self-ratings of adolescent personality in 24 cultures (McCrae et al., 2010). Foreshadowing the current investigation, Putnam and Gartstein (2017) correlated nation-level personality variables reported by McCrae (2001) and McCrae et al. (2005) with the aggregate temperament scores from their meta-analyses, finding countries high on adult Extraversion were similarly high on child Surgency, those high on self-reported Neuroticism were high on Negative Affectivity, and country-level other-reported Agreeableness correlated positively with both Surgency and RC. In the current study, we replicate and extend these analyses.

Our fifth question may be most informative in terms of beginning to understand the societal values that may form the bases for different developmental niches inhabited by children around the world: How is temperament associated with dimensions of cultural orientation and national wealth? Aspects of cultural philosophy are frequently proposed as causal factors explaining differences between citizens of different nations. These arguments, however, typically rest upon anecdote and speculative logic, rather than in quantitative analyses. That is, differences observed between samples from a few locations are credited to specific aspects of culture, but the nations under investigation differ in myriad ways other than the dimensions suggested. Examples are common in the literature relating culture and temperament, with the distinction between individualist and collectivist orientations frequently proposed as contributing to differences. In comparing Canadian and Chinese children with respect to shyness, Chen (2000) published the paper “Growing up in a collectivist culture: Socialization and socioemotional development in Chinese children”. In their examination of four cultures, Gartstein et al. (2010) wrote “…selection of these countries (Japan, the U.S., Poland and Russia) presented an opportunity to conduct comparisons between cultures that vary on the individualistic/collectivistic value systems.” More recently, Krassner et al. (2017) authored “East-west, collectivist-individualist: A cross-cultural examination of temperament in toddlers from Chile, Poland, South Korea, and the U.S.”. Although it is plausible that differences between cultures in the degree to which they promote an individualist focus on the self versus the collectivist importance of the larger group may influence displays of temperament, it is also the case that Canada and China, Russia and

Japan, and Chile and Poland vary widely in other aspects of cultural practices and beliefs. Only by studying individualism-collectivism and temperament across a large number of nations can one achieve the statistical power necessary to confirm this relation.

Other dimensions of cultural orientation have been less regularly applied to cross-cultural differences in psychological constructs, including temperament. The most frequently used framework for characterizing cultural dimensions is that derived by Geert Hofstede and colleagues (1984, 2001, 2011; Hofstede et al., 2010; Minkov, 2007). While recognizing that culture was manifest in multiple levels, with regional, ethnic, social class and work organizations exhibiting their own cultural proclivities, Hofstede (e.g., Hofstede et al., 2010) argued that there was substantial value to studying national culture differences, as forces such as national educational systems, financial markets and political bodies compel individuals to integrate within the dominant national structure. Furthermore, research on cultural differences is typically conducted on the national level, as use of this distinction is more expedient than gathering data from more nuanced self-organizing societies. In addition to Individualism/Collectivism, analyses by Hofstede and his collaborators suggested five additional, relatively independent aspects of national culture: Power Distance (acceptance of inequality in power in a society), Masculinity/Femininity (the extent to which a society is driven by competition, achievement, and success, rather than cooperation, modesty, nurturance, and a focus on consensus), Uncertainty Avoidance (the extent to which society members are threatened by unstructured situations that are novel, unknown, surprising, or unusual), Long-Term/Short-Term Orientation (emphasis on values of persistence, thrift, and having a sense of shame versus reciprocating favors and protecting one’s ‘face’ to satisfy more immediate desires), and Indulgence/Restraint (the degree to which a given society allows members to be unrestrained to pursue hedonic pursuits).

Despite their differing theoretical bases in defining groups versus describing individuals, substantial links between Hofstede’s dimensions and national-level personality/temperament have been uncovered. For instance, Hofstede and McCrae (2004) and Bartram (2013) found Individualism to be strongly associated with countries’ average scores on Extraversion, while a large amount of variance in Neuroticism was explained by cultural Uncertainty Avoidance and Masculinity. Using meta-analytic data from 18 countries, Putnam and Gartstein (2017) identified multiple correlations between aggregate temperament scores and cultural orientation. In their analyses, high Surgency was associated with low Power Distance, Short-Term Orientation and Indulgence; high Negative Affectivity with Collectivism, high Power Distance, Masculinity, high Uncertainty Avoidance and Restraint; and high Regulatory Capacity with low Power Distance and Femininity. A replication attempt with the smaller (N=14) JETTC dataset (Gartstein & Putnam, 2018) confirmed only the relation between Collectivism and Negative Affectivity, but nonsignificant trends were in the same direction for eight of the nine associations (not Masculinity-Negative Affectivity) that had been significant in the meta-analysis. The large number of nations included in the GTP provide an opportunity for a more thorough assessment of the relations between cultural orientation and early temperament.

A final variable considered in relation to nation-level estimates of temperament is the relative wealth of the countries’ citizens. Several studies have shown national income to be higher in countries that are more Individualistic and low in Power Distance (Hofstede et al., 2010). Moreover, within Australia, Strickhouser and Sutin (2020) found infants developing in lower SES households demonstrated lower sociability (analogous to Surgency), higher reactivity (analogous to Negative Affectivity) and lower persistence (analogous to Regulatory Capacity), and Parade & Leerkes (2008) found family income to predict higher parent ratings of High

Intensity Pleasure, Approach, Perceptual Sensitivity and Distress to limitations in US infants. Putnam & Gartstein (2017) extended this research to nation-level differences, finding high Gross National Product (GNP) per capita to be associated with lower Negative Affectivity and higher Regulatory Capacity. We anticipate similar findings in the GTP.

Our sixth and final research question is perhaps the most intuitive: What are the geographical patterns of temperament traits? In contrast to early studies involving few countries, inclusion of samples from a larger number of locations allows for exploration of patterns across widespread world areas. Previous studies of personality and temperament distribution inform our expectations of geographically proximal countries. Analyses of adult personality by McCrae and Allik (Allik et al., 2017; Allik & McCrae, 2004; McCrae, 2001; McCrae, 2002; McCrae et al., 2005) suggested that samples from Europe and the United States tended to demonstrate higher Extraversion and Openness than Asian and African cultures; and those from Northern European nations evinced higher Neuroticism than Southern European countries. With respect to temperament, Putnam and Gartstein (2017) reported consistently low Surgency, high Negative Affectivity, and mostly low Regulatory Capacity in East Asian nations; high levels of Negative Affectivity among countries in Eastern Europe; and low Negative Affectivity and high Regulatory Capacity in Northern European cultures. Results from the JETTC (Gartstein & Putnam, 2018) were roughly consistent with these trends. Again, Negative Affectivity was high in toddlers from East Asia, as well as Türkiye, and low among children from Northern Europe, and from Italy. Negative Affectivity scores also differed between Latin American countries, with significantly higher scores in Brazilian and Chilean than Mexican samples, similar to results for adult Neuroticism (McCrae et al., 2005). Consistent with Putnam & Gartstein (2017), Surgency was low in several East Asian cultures. In addition, toddlers from

Finland, Belgium, and Chile were rated high in Surgency. JETTC findings regarding Regulatory Capacity differed somewhat from the 2017 meta-analyses. Although toddlers from China were rated low on Regulatory Capacity, those from South Korea received relatively high scores. Results among European nations were also inconsistent with the meta-analyses, as Northern European children were not rated particularly highly in Regulatory Capacity. In the current study, we anticipate confirming the consistent findings of high Negative Affectivity and low Surgency in most, if not all, East Asian nations, and low Negative Affectivity in Northern Europe; and resolve inconsistencies regarding patterns of Regulatory Capacity in the two previous multicultural investigations.

Summary of Goals

The current collaboration provides a more powerful lens than prior efforts to answer the following questions regarding the role of culture in shaping early emerging individual differences in reactivity and regulation: (1) What is the size of national-level differences in temperament at different ages and for different dimensions? (2) To what degree are gender differences in temperament consistent or inconsistent across nations? (3) Are patterns of nation-level differences in temperament consistent across infancy, the toddler period, and early childhood? (4) Are nation-level differences in temperament consistent with previous studies of nation-level personality? (5) How are national temperament aggregate scores associated with dimensions of cultural orientation and national wealth? (6) How are temperament traits distributed geographically?

Methods

Samples

The majority of samples were obtained through an email outreach effort. All researchers from outside the United States (US) who requested access to any version of the IBQ-R, ECBQ and/or CBQ between September 2006 and November 2017 through the Rothbart Temperament Questionnaires website (https://research.bowdoin.edu/rothbart-temperament-questionnaires/) were contacted with a request to share their item-level temperament data, in addition to age (at time of collection) and gender information for each subject. Approximately 2000 researchers were contacted in this way, and approximately 275 datasets were ultimately shared on the basis of this contact. In addition, all 16 investigators from the JETTC were contacted and agreed to share their data with the GTP. Similarly, researchers whose infant or child data were used in the Putnam and Gartstein (2017) meta-analyses, and those whose data were used in the construction of short forms of the instruments, were contacted with requests to share their item-level data. Finally, a small number of datasets were contributed by researchers who learned about the GTP through word of mouth. These researchers used a variety of procedures to recruit participants and collect data. Information regarding these details can be found in supplementary file 1, available online. Collectively, 377 independent datasets were compiled. Appendix 1 indicates the names and institutional affiliations of all GTP partners who collaborated in the individual studies from which these data were drawn.

Of the 377 datasets acquired, 36 were not included in our analyses, for a variety of reasons. Twelve datasets exhibited very poor internal consistency estimates (i.e., alphas < .50 for two of the three scales). Three datasets were collected using the original (Rothbart, 1981) IBQ, rather than the IBQ-R. Eight datasets contained data missing several items from one or more scales. Four datasets used unconventional scoring response options (e.g., items rated on scales ranging from 1-5, rather than 1-7). Two were collected using teacher-report, rather than parent-report, forms. One dataset contained no age or gender data. Three contained scale-level, rather than item-level, data. Finally, three datasets contained fewer than 25 cases. From the remaining 341 datasets, 17 were flagged for potentially unrepresentative samples. For five of these, participants were selected in relation to birth factors (e.g., premature delivery, low birth weight). Five had been recruited on the basis of child factors (e.g., speech delays, anxiety symptomatology) that may affect temperament. Two were selected for maternal characteristics (e.g., high stress, depression). Five were selected for demographic status (e.g., socially disadvantaged areas). Although these child, maternal and demographic characteristics do not necessarily render the samples unrepresentative of the larger populations from which they were drawn, they potentially represent different segments of their areas than other samples, which were largely community/convenience samples. In addition, datasets were examined with respect to representativeness. Specifically, we identified 14 samples with average scores for at least one temperament score that was more than one SD different from the average scores for the other samples from their country. The analyses reported were conducted when datasets were removed for each of these five reasons, and substantive findings were unchanged. Therefore, the flagged samples were included in the presented results.

Of the 341 datasets used in the analyses, data collection for 36 was longitudinal across spans covered by more than one questionnaire (e.g., the IBQ-R and ECBQ). For these samples, data from both time points were used in analysis involving individual questionnaires and in creation of scores combined across questionnaires. In addition, for several datasets, longitudinal data were collected using the same questionnaire (e.g., the IBQ-R at 3, 6 and 9 months). For these datasets, a set was created in which each child was represented by data collected at a single time point.

For the large majority of samples, data were collected through primary caregivers' completion of the standard (full), short, or very short forms of the IBQ-R, ECBQ, and/or CBQ. For a handful of samples, researchers used a customized version of the instrument (e.g., all very short form items, plus full scales for select dimensions). Another small group used different forms (e.g., the standard and short CBQ) for subsets of their sample. For the large majority of samples, parents reported the age and gender of their child. For a handful of samples, medical records confirmed or were used in the place of parent report of gender and age. Information regarding all samples, including the community from which they were collected, the form used, the number of participants for each questionnaire, and whether the dataset was flagged for relevant maternal, child or demographic characteristics can be found in supplemental file 1, available online.

The IBQ-R was originally developed through analyses of infants between 3 and 12 months of age, the ECBQ was developed using data for children between 18 and 36 months, and the CBQ with children from 3 to 8 years. Subsequent research, including that concerning the development of the abbreviated measures (Putnam et al., 2014; Putnam & Rothbart, 2006), has indicated successful use of the Very Short Forms in children younger or older than the intended range. In the GTP data, IBQ-Rs were completed by parents of infants ranging from 0 to 18 months; ECBQs were completed for children ages 12 to 60 months; and CBQs were completed for children ages 24 to 120 months. To address age differences between samples, age was entered in a covariate in all substantive analyses.

Measures

The IBQ-R, ECBQ, and CBQ are parent-report instruments containing items referring to infant and child behavior in commonly occurring situations. For each item, parents are asked to rate the child on a 7-point Likert-type scale. The IBQ-R and ECBQ items are phrased in the form of questions about the child's behavior in a given context during the past 1 or 2 weeks (e.g., "When being carried in the past week, how often did the baby push against you until put down?"), and the ratings refer to frequency of behavior (never, very rarely, less than half the time, half the time, more than half the time, almost always, always). The CBQ items are statements describing child behavior (e.g., "My child gets angry when told he or she needs to go to bed"), and the ratings refer to the degree to which the statement accurately describes the child's behavior in the past 6 months (extremely untrue, quite untrue, slightly untrue, neither true nor untrue, slightly true, quite true, extremely true). For all items on the three questionnaires, parents are also given the option of choosing "NA" if they have never observed their child in the situation described.

The standard forms of the IBQ-R, ECBQ, and CBQ contain 191, 205 and 195 items, respectively, and measure between 14 and 18 fine-grained scales (see Gartstein and Rothbart, 2003; Putnam et al., 2006 and Rothbart et al., 2001 for details). Short forms (91, 107 and 94 items) were subsequently developed which used fewer items from the original forms to measure the same scales, as well as very short forms consisting of 12- or 13-item scales intended to measure only the three broad factors of Surgency, Negative Affectivity and Regulatory Capacity (see Putnam et al., 2014; Putnam et al., 2010; and Putnam & Rothbart, 2006 for details). Since the development of the original forms in English, the psychometric characteristics of translations of the measures across multiple languages have been described in several publications (e.g., Barcenilla, Luttges, Rojas-Barahona, & Campos, 2021; Costa & Figueiredo, 2018; Golmohammadi et al., 2020; Sleddens et al., 2011; Stępień-Nycz et al., 2018).

Data Management

To ensure that all scales were calculated consistently across datasets, investigators were asked to send files containing item-level data, in addition to child age and gender. An initial step was to transfer data into common SPSS templates for the standard, short or very short forms of the IBQ-R, EBCQ and CBQ. All cases collected from children far outside the recommended ages for the three instruments (i.e., IBQ-Rs older than 18 months; ECBQs younger than 12 months or older than 60 months; CBQs younger than 24 months or older than 120 months), as were all cases with missing data for all questionnaire items. Standard and short form data of the IBQ-R, ECBQ and CBQ were transformed into very short form data files by selecting out only the items from the longer forms that were included in the very short measures and merged into a single dataset. Missing data for at least one item was present for 86% of IBQ-R cases, 49% of ECBQ cases, and 25% of CBQ cases. These missing items were replaced using maximum likelihood estimation. Scale scores were calculated as the average of item scores corresponding to each scale.

Testing Measurement Invariance

Tests of measurement invariance were guided by Leerkes et al. (2017), Byrne and van de Vijver (2010), and Senese et al. (2012). Leerkes et al. (2017) established the measurement invariance of the IBQ-R-VSF across U.S. samples differing in race and poverty status. Two elements of their analyses warrant special attention. First, upon observing poor fit in Confirmatory Factor Analysis (CFA) that assumed all error terms were uncorrelated, they followed the approach used by Putnam and Rothbart (2006) in their initial evaluation of the CBQ-VSF by allowing a-priori correlations between error terms from items taken from the same subscales. For example, items from the IBQ-R VSF SUR scale were taken from longer scales measuring Activity Level, Smiling and Laughter, High Intensity Pleasure, Perceptual Sensitivity and Approach; and the modified CFA model allowed correlated errors for items from Activity Level, Smiling and Laughter, etc. Although modification indices suggested additional correlated errors, Leerkes et al. (2017) did not allow additional correlations on these bases. As previous research (e.g., Rothbart et al., 2001; Gartstein & Rothbart, 2003) typically demonstrates correlations between factors, these were allowed in the base and error-correlated models. Second, given arguments that personality measures often demonstrate poor CFA fit due to the complex nature of personality (Hopwood & Donnellan, 2010) and that criteria for traditional fit estimations (e.g., comparative fit indices [CFI] over .90) are excessively restrictive when applied to measures with multiple items (Marsh et al., 2004), Leerkes et al. (2017) referenced Kenny (2014) in considering CFI > .85 and root mean square error of approximation (RMSEA) < .05 as demonstrating acceptable fit.

Responding to the difficulty in interpreting model fit in large-scale cross-cultural studies, as problematic results may either be due to properties of the instrument or issues concerning data from individual countries, Byrne and van de Vijver (2010) proposed and exemplified a two-pronged approach when confronted with evidence of measurement inequivalence among multiple samples. In the first step, descriptive statistics and factor loadings for individual countries and items are evaluated as potential contributors to poor fit. In the second, a series of CFA models are tested to identify sources of compromised fit. Following both Byrne and van de Vijver (2010) and Leerkes et al. (2017), we considered Δ CFI > .01 to indicate a significant change in fit (see Cheung & Rensvold, 2002).

Of more critical importance than the fit of the overall model is the equivalence of measurement within each scale. As such, although the analyses presented below indicate a lack of configural variance for the full questionnaires, consistent with Senese et al. (2012), who separately evaluated the cross-cultural equivalence of separate scales comprising a multi-factor measure of parenting beliefs, efforts were taken to derive scales for SUR, NEG and EFF that measured these constructs equally across nations. To do so, configural models for each of the three scales were evaluated with the 20 largest nationwide samples. To explore potential improvements to fit for these scales, we referred to Byrne and van de Vijver's (2010) recommendations to identify potentially problematic items and countries. First, country-level means for all items were examined for outliers. Following this, to determine whether problems with fit were due to measurement in certain countries, a series of models was run in which countries containing outliers on at least one variable were either eliminated if they were among the 20 largest samples, or added if they were not among the 20 largest. Next, a series of CFAs were conducted in which each individual item was removed from the configural model for each factor to determine whether omitting the item improved model fit. Our application of this approach is described below. CFA was conducted using SPSS AMOS 27.

IBQ-R-VSF

First, the base model with uncorrelated errors was tested with the full sample of IBQ-R VSF data. As expected, the fit of this model was poor, CFI = .619, RMSEA = .078. Allowing a-priori correlated errors improved fit to an acceptable level, CFI = .859, RMSEA = .049. Next, the configural model was computed to examine whether the hypothesized factor structure was supported in the 20 countries with the largest sample sizes (ns > 250). Fit of this model was unacceptable, CFI = .762, RMSEA = .014.

When the SUR scale with errors correlated was evaluated alone using the entire dataset, the fit was acceptable, CFI = .930, RMSEA = .060. When the configural model was tested with the 20 largest nation samples, fit diminished substantially, CFI = .834. RMSEA = .018. During examination of national means, five outliers were identified. Scores for Iran were discrepant for items 2, 8 and 14. Latvia and Malaysia were outliers for item 26. Configural model fit did not change (i.e., ΔCFI < .01) when Latvia was removed or Malaysia or Iran were added. Examination of CFI change upon removal of items indicated that both item 8 and item 26 improved the fit of the configural model CFI > .01 when eliminated. When both items 8 and 26 were removed, CFI = .866, RMSEA = .017. Item 8 (laughs when put in bath) represents the Smiling and Laughter scale in the standard IBQ-R, while item 26 (vocalizes when hair is washed) represents Vocal Reactivity. The contribution of these items to misfit suggests different bathing contexts across countries.

When the NEG scale was evaluated, fit was acceptable, CFI = .975, RMSEA = .047. When the configural model was tested with the 20 largest nation samples, fit diminished to CFI = .871, RMSEA = .018. National mean examination indicated discrepancies for Belgium (item 3), Nigeria (item 17), Kosovo (item 28), Malaysia (item 28), Chile (item 32) and the Netherlands (item 32). Removal or addition of countries did not impact configural model fit. When items were removed, items 9, 17 and 33 each resulted in CFI improvement > .01. Following Byrne & de Vijver (2010), we then investigated which combination of two or three of these items led to the maximum model fit when eliminated. Elimination of all three items resulted in the best fit, CFI = .953, RMSEA = .012. Item 9 (whimpers and cries when time for bed) represents Sadness in the standard IBQ-R, item 17 (startles at change in body position) represents Fear, and item 33 (clings to parent in presence of unfamiliar adults) also represents Fear. The best fitting model thus includes compromised representation of Fearfulness.

When REG was evaluated, fit was acceptable, CFI = .915, RMSEA = .073. When the configural model was tested with the 20 largest nation samples, CFI = .808, RMSEA = .20.

National mean outliers were apparent for Romania (item 18), Poland (item 18), Malta (item 25), Sweden (item 31) and China (34). Removal of China resulted in CFI = .826, RMSEA = .019. When items were removed individually, items 11, 19, 30 and 34 improved CFI > .01. We then investigated the combination of items leading to maximum fit, in models both containing and eliminating China. When China was included in these analyses, a model excluding items 30 and 34 resulted in CFI = .875, RMSEA = .018 and removing additional items did not improve fit. However, with China excluded, CFI = .892 with 30 and 34 removed, but rose to .907 when all four items were taken out, suggesting that items 11 and 19 were particularly problematic in the Chinese sample. Because we wished to retain the maximum number of items to maintain content validity, we considered the fit with items 30 and 34 out as the most useful model. Item 30 (enjoys gentle rhythmic activities such as rocking or swaying) represents Low Intensity Pleasure and item 34 (enjoys when rocked or hugged) represents Cuddliness. The contribution of these items to misfit suggests different meanings of these descriptions, or differing contexts of parents rocking their infants across countries.

Across the entire sample, alphas for the original IBQ-R SUR, NEG and REG scales = .81, .80 and .75, respectively. In the 116 individual datasets (see supplemental data available online), alphas for SUR ranged from .47 to .89 with alphas < .60 for 8 samples; alphas for NEG ranged from .60 to .89; and alphas for REG ranged from .49 to .85 with alphas < .60 for 4 samples. When revised scales were calculated to reflect the best fitting models from our configural CFA analyses, alphas for the entire sample for SUR, NEG and REG = .79, .76 and .73. In the 116 individual datasets, alphas for the abbreviated SUR scale ranged from .40 to .86 with alphas < .60 for 11 samples; alphas for abbreviated NEG ranged from .55 to .85 with alphas

< .60 for 5 samples; and alphas for abbreviated REG ranged from .51 to .83 with alphas < .60 for 8 samples.

ECBQ-VSF

First, the base model with uncorrelated errors was tested with the full sample of ECBQ data. As expected, the fit of this model was very poor, CFI = .598, RMSEA = .066. Allowing a-priori correlated errors improved fit, although fit remained poor, CFI = .716, RMSEA = .057. Fit diminished further when the configural model was tested with the 20 largest samples (ns > 189), CFI = 657, RMSEA = .014.

When the SUR scale was evaluated alone with the entire sample, the fit was acceptable, CFI = .862, RMSEA = .070. When the configural model was tested with the 20 largest nation samples, fit diminished slightly, CFI = .855. RMSEA = .017. Regardless, we followed steps to find a best-fitting model across nations. During examination of national means, outliers were apparent for China (items 18 and 30), Colombia (items 2 and 4), Mexico (items 9 and 25), Nigeria (item 20), Taiwan (item 13) and Thailand (item 35). Model fit did not change (i.e., Δ CFI < .01) when China or Taiwan were removed or the smaller nations were added. Examination of CFI change upon removal of items indicated that five items improved CFI > .01 when eliminated. Removal of item 9 resulted in the greatest change, CFI = .883. Consistent with Byrne & de Vijver (2010), we explored combinations of other items removed with item 9 to identify a best-fitting model. Because a model eliminating items 9 (becomes excited when loved ones are to visit), 13 (gets involved immediately in new activities) and 11 (likes rough and rowdy games), CFI = .927 was not improved > .01 with elimination of additional items, it was retained as the final model. Because these items each come from different scales in the standard ECBQ, the resulting abbreviated scale seems to retain the content of the original VSF SUR scale.

When the NEG scale was evaluated alone with the entire sample, the fit was acceptable, CFI = .890, RMSEA = .069. When the configural model was tested with the 20 largest nation samples, CFI = .863, RMSEA = .017. Again, we followed steps to find a best-fitting model across nations. Examination of national means revealed outliers for Brazil (item 22), Chile (item 23), China (item 23), Colombia (items 1, 2, 22, 26, and 33), Kosovo (items 16, 17, 19, 22 and 23), Lithuania (items 2 and 22), Nigeria (item 19), Portugal (item 22), Sweden (item 22), and Switzerland (item 22). CFI did not change > .01 when these nations were added or removed. Examination of CFI change upon removal of items identified four items that improved CFI > .01 when eliminated. We explored combinations of elimination for the four items. Removal of items 26 (has a temper tantrum when told no) and 17 (bothered by noisy environments) resulted in CFI = .940. Because elimination of additional items did not increase CFI > .01, this was kept as the final model. Poor cross-cultural fit of these items suggest difficulty in translating the concept of a tantrum and environmental differences between nations.

When the REG scale was evaluated alone with the entire sample, fit was unacceptable, CFI = .827, RMSEA = .079. When the configural model was tested with the 20 largest nation samples, CFI = .797, RMSEA = .019. Steps were taken to find a best-fitting model. Examination of nation means revealed outliers for Australia (item 15), Colombia (item 31), France (item 7), Germany (item 8), Japan (item 5), Kosovo (item 27) and Mexico (item 23). CFI did not change > .01 when Germany or Japan were removed or the smaller nations were added. When individual items were removed, seven items improved fit CFI > .01. Elimination of the two items contributing most to misfit (28, smiles when rocked; and 7, plays with toy for more than 10 minutes) improved fit to CFI = .875. When the other five items were removed one-by-one, removal of item 14 (tires of activities requiring attention) resulted in the greatest improvement, CFI = .911. The remaining four were removed one-by-one, with removal of item 15 (pays attention right away when called) resulting in CFI = .936. No remaining deletions resulted in substantial increases in fit. These removals included both attention focusing items, one attentional shifting and one cuddliness item, thus resulting in a scale with less emphasis on attention control than the original.

Across the entire sample, alphas for the original ECBQ SUR, NEG and REG scales = .71, .75 and .74, respectively. In the 99 individual datasets (see supplemental data available online), alphas for SUR ranged from .46 to .82 with alphas < .60 for 6 samples; alphas for NEG ranged from .46 to .82 with alphas < .60 for 11 samples; and alphas for REG ranged from .52 to .86 with alphas < .60 for 4 samples. When revised scales were calculated to reflect the best fitting models from our CFA analyses, alphas for the entire sample for SUR, NEG and REG = .66, .69 and .65. In the 116 individual datasets, alphas for the abbreviated SUR scale ranged from .32 to .80 with alphas < .60 for 22 samples; alphas for abbreviated NEG ranged from .36 to .79 with alphas < .60 for 44 samples; and alphas for abbreviated REG ranged from .38 to .81 with alphas < .60 for 39 samples. As such, elimination of items to enhance cross-cultural comparability took a substantial toll on the internal consistency of scales.

CBQ-VSF

First, the base model with uncorrelated errors was tested with the full sample of CBQ data. As expected, the fit of this model was very poor, CFI = .566, RMSEA = .072. Allowing a-priori correlated errors improved fit, although fit remained unacceptable, CFI = .799, RMSEA = .05. When the configural model was tested with the 20 largest nation samples (ns >310), fit diminished modestly, CFI = .776, RMSEA = .013.

When the SUR scale was evaluated alone with the entire sample, the fit was unacceptable, CFI = .806, RMSEA = .095. When the configural model was tested with the 20 largest nation samples, fit increased but not to acceptable levels, CFI = .816, RMSEA = .022. We then followed steps to find a best-fitting model across nations. During examination of national means, outliers were apparent for Lithuania (items 10 and 31), Malaysia (item 13, Myanmar (item 1), Pakistan (item 1) and Thailand. Adding these nations to the analyses did not diminish fit. Examination of CFI change upon removal of items indicated that 6 items improved CFI > .01 when eliminated. Removal of item 19 (takes time approaching new situations) dramatically improved fit (Δ CFI = .072; CFI = .888). Additional removal of item 7 (often rushes into new situations) improved fit greatly (Δ CFI = .044; CFI = .932), and removal of item 31 (unhurried in deciding what to do next) after this resulted in CFI = .960. No further refinements were attempted. The three problematic items were all from the Impulsivity scale from the original CBQ, suggesting this factor does not equally contribute to SUR across nations.

When NEG was evaluated with the entire sample, the fit was good, CFI = .955, RMSEA = .046. When the configural model was tested with the largest samples, fit did not diminish substantially, CFI = .948, RMSEA = .011. Regardless, potential improvements to fit were examined. Nations with outlying item scores included Colombia (item 23), Japan (item 32), Malaysia (item 11), Myanmar (items 17 and 20), the Netherlands (item 35) and Türkiye (item 17). Exclusion of Türkiye, Japan or the Netherlands from the configural model did not alter CFI > .01. Fit was then evaluated with individual items removed. Although removal of item 2 increased CFI to .960, the original scale was retained to maintain content validity.

When REG was evaluated with the entire sample, fit was good, CFI = .970, RMSEA = .039. When the configural model was tested with the largest samples, fit diminished to CFI =

.956, RMSEA = .011. Nations with outlying item scores included Brazil (item 24), Colombia (items 3, 6, 12, 24, and 36), Greece (item 23), India (item 30), the Netherlands (item 6), Romania (items 33 and 36) and Thailand (item 9). Removal of Greece or the Netherlands or addition of the smaller countries did not diminish fit. Fit was then evaluated with individual items removed. Although removal of item 27 increased CFI to .970, the original scale was retained to maintain content validity.

Across the entire sample, alphas for the original CBQ SUR, NEG and REG scales = .71, .75 and .75, respectively. In the 172 individual datasets (see supplemental data available online), alphas for SUR ranged from .15 to .88 with alphas < .60 for 25 samples; alphas for NEG ranged from .436 to .82 with alphas < .60 for 14 samples; and alphas for EFF ranged from .46 to .89 with alphas < .60 for 7 samples. When the revised scale for SUR was calculated to reflect the best fitting model from our CFA analyses, alpha for the entire sample = .63. In the individual dataset, alphas for the abbreviated SUR scale ranged from .08 to .87 with alphas < .60 for 57 samples. As with the ECBQ scales, shortening this scale dramatically impacted internal consistency.

Archival Data

Aggregate self-report personality scores were obtained from Allik et al. (2017) and comprised average scores for college students and adults in the given countries on the five NEO Personality Inventory factors. McCrae (2001, 2002) initially obtained data for 36 countries from researchers who had collected samples for their own studies. Although the exact years of data collection for these samples were not reported, the publication dates for articles stemming from these data ranged from 1992 to 2001. Allik et al. (2017), supplemented these samples with new entries, including those from reports published between 2004 and 2014, to arrive at a total of 76 samples for 62 countries. For countries with multiple samples, sample scores were averaged to create the scores used in the current analyses. Aggregate other-report personality scores for 51 countries were obtained from McCrae et al. (2005a). The majority of these samples were gathered from college students asked to rate either a college-aged or an adult individual from their country that they knew well, while three samples comprised existing datasets for which raters were spouses or peers (McCrae et al., 2005b). Although McCrae et al. (2005a, 2005b) did not indicate the years during which these data were collected, language in their report framing this study as a follow-up to McCrae (2001, 2002) suggests they were gathered between 2001 and 2005. Of the 59 nations represented in the GTP, aggregates of self-reported personality were available for 37 (not for Argentina, Bosnia-Herzegovina, Brazil, Chile, Colombia, Curacao, Iran, Ireland, Israel, Kosovo, Malta, Myanmar, Nigeria, Pakistan, Servia, Singapore, Slovakia, Slovenia, Suriname, Thailand, Ukraine or Uruguay), and other-reported personality aggregates were available for 37 (not for Bosnia-Herzegovina, Colombia, Curacao, Finland, Greece, Hungary, Ireland, Israel, Kosovo, Latvia, Lithuania, Myanmar, Netherlands, Pakistan, Philippines, Romania, Singapore, Suriname, Sweden, Taiwan, Ukraine, or Uruguay).

Values for Hofstede’s six dimensions, initially published by Hofstede (2001), Hofstede and Hofstede (2005), and Hofstede et al. (2010) were obtained from http://www.geerthofstede.nl/research--vsm on July 22, 2015. The initial four dimensions in Hofstede’s system (Individualism-Collectivism, Masculinity-Femininity, Uncertainty Avoidance, and Power Distance), were developed from questionnaires administered to IBM employees in 71 nations between 1967 and 1973. A fifth dimension, Long Term Orientation-Short Term Orientation, emerged from analyses in the 1980s of an instrument initially designed to assess basic values of Chinese citizens and subsequently administered in multiple nations (Hofstede &

Bond, 1988; Hofstede et al., 2010). The sixth dimension, Indulgence-Restraint, was developed in the 2000s by Misho Minkov and incorporated into Hofstede's work (Hofstede et al, 2010). The majority of scores published in Hofstede's publications (e.g., Hofstede et al., 2010) and used in the current paper were collected during the initial data collection of the first four dimensions in 1969, and in the early 2000s for the two dimensions added later (Geert Jan Hofstede, personal communication, June 22, 2023). Recent research (Beugelskijk et al. 2015) indicates that, although scores on Hofstede's dimensions have demonstrated absolute change (i.e., more recent birth cohorts exhibit higher Individualism and Indulgence, but lower Power Distance than previous cohorts), relative scores of countries exhibit little change, suggesting their continued validity.

Scores for all cultural dimensions were available for 51 of the 59 countries represented in our analyses. No cultural dimension scores were available for Curacao, Kosovo or Myanmar; scores for Long Term Orientation-Short Term Orientation were not available for Suriname; scores for Indulgence-Restraint were not available for Suriname and Israel; and scores for Individualism-Collectivism, Masculinity-Femininity, Uncertainty Avoidance, and Power Distance were not available for Nigeria, Bosnia-Herzegovina and Ukraine.

Following Hofstede and McCrae (2004) and Putnam and Gartstein (2017), who included Gross National Income Per Capita (GNI-PC) as a control variable when exploring relations between aggregate personality/temperament and cultural orientation, these values were obtained from http://data.worldbank.org/indicator/NY.GNP.PCAP.CD for 58 countries. For 55 nations, a single estimate was created by averaging GNI-PC from years 2000-2016. For Curacao, Greece and Kosovo, GNI-PC values were only available for more recent years. For these three nations, values comparable to the 2000-2016 average were created by regressing 2000-2016 estimates on averages for the available years and extrapolating. No GNI-PC scores were available for Taiwan, and an estimate for Taiwan was created by extrapolating from 2019 Gross National Product data obtained from https://en.wikipedia.org/wiki/List_of_countries_by_GDP_(nominal)_per_capita.

To examine larger global patterns of temperament, all nations were organized with respect to the geographic subregions used by the Statistics Division of the United Nations (obtained from https://unstats.un.org/unsd/methodology/m49/). Countries were classified within 14 regions as follows: Australia and New Zealand, Caribbean (Curacao), Central America (Mexico), Eastern Asia (China, Hong Kong, Japan, South Korea, Taiwan), Eastern Europe (Czech Republic, Hungary, Poland, Romania, Russia, Slovakia, Ukraine), Northern America (Canada, United States), Northern Europe (Denmark, Estonia, Finland, Ireland, Latvia, Lithuania, Sweden, United Kingdom), South America (Argentina, Brazil, Chile, Colombia, Peru, Suriname, Uruguay), Southeastern Asia (Indonesia, Malaysia, Myanmar, Philippines, Singapore, Thailand), Southern Asia (India, Iran, Pakistan), Southern Europe (Bosnia and Herzegovina, Greece, Italy, Kosovo, Malta, Portugal, Serbia, Slovenia, Spain), Sub-Saharan Africa (Nigeria), Western Asia (Israel, Türkiye), and Western Europe (Belgium, France, Germany, The Netherlands, Switzerland).

Due to intellectual property agreements with participating sites, raw data are not publicly available, although summary files will be made available upon request to the corresponding author.

Results

Analyses addressing our six research questions were conducted using the revised scale scores emerging from our investigation of measurement invariance, for which all items functioned relatively consistently across nations.

Although Multilevel Modeling (MLM) is frequently used to account for nested data (e.g., collection sites as Level 2 within nation as Level 3), this approach was unacceptable with the GTP data, as 21 of the 59 nations were represented by only one dataset. Similarly, the number of larger geographical regions (N = 14) was not large enough to have adequate statistical power and stable estimates, and some regions were represented by a single nation (see supplemental file 1, available online). For instance, with fewer than 30 groups, standard errors tend to be too small, inflating Type 1 error. Recent guidelines (Hox & McNeish, 2020) with estimation procedures used in this study suggest more than 20 groups for accurate fixed effect estimates at the higher levels. MLM analyses were conducted with cases nested at the levels of nation and global region separately (see supplemental file 2, available online). Because these findings were largely similar to those reported below with more conventional statistics, to enhance comparability between our findings and those obtained in previous studies, our presented analyses largely mirror those employed by others (e.g., Allik et al., 2017; Putnam & Gartstein, 2017).

In addition to proportional analyses mirroring Allik et al. (2017), ANOVA were used to address questions 1 and 2 regarding the size and nature of nation and gender effects on temperament scores from the three questionnaires. The marginal means generated by these ANOVA formed the basis for correlational analyses of cross-age consistency (question 3). To maximize sample size, reduce the number of tests, and enhance interpretability regarding geographical patterns and relations with aggregate personality and cultural orientation (questions 4, 5 and 6), for each temperament dimension a single “omnibus” value for each country was created by standardizing the marginal mean scores from the three different measures and averaging these scores. Questions 4 and 5 were addressed through correlations between these omnibus scores and archival nation-level scores for adult personality and Hofstede’s cultural orientation scores. The descriptive goal of Question 6 regarding geographical distribution of temperament traits was addressed through maps for which omnibus scores were represented by differential shading. Interpretation of these geographical patterns was further aided by analyses of questionnaire scores at the UN region level.

Question 1: Size of Nation-Level Effects

The magnitude of relations between temperament and culture was assessed in two ways. First, following Allik et al. (2017), the standard deviations of the mean values of the nine dimension scores (Surgency, Negative Affectivity and Regulatory Capacity for the IBQ-R, ECBQ and CBQ) were calculated for each nation, and these within-nation standard deviations were averaged across all nations. Next, the standard deviation of the mean values from these nations (i.e., between-nation standard deviations) was calculated. These standard variations were then squared to reveal variance, and the proportion of between- to within-nation variance calculated. As reported in Table 1, in comparison to Allik et al. (2017), who reported average proportions of 11.8 for adult personality scores, the average proportion was 14.0 across the 9 temperament scores, ranging from 4.9 to 20.0.

Magnitude of nation-level effects were also explored through Nation by Gender ANOVAs, with Age as a covariate. Results, shown in Table 1, indicate significant effects of Nation for all temperament dimensions. Variance accounted for by Nation ranged from .034 to .168 (average = .078). In comparison, the amount of variance explained by Gender ranged from .000 to .009 (average = .002).

The relative amount of variance explained by nation across the nine dimensions was largely consistent across the two analyses. The effects of nation were most pronounced for Negative Affectivity, particularly on the ECBQ. Substantial effect of nation were also apparent for Regulatory Capacity. ANOVAs indicated the smallest effects of nation for Surgency across all three measures.

Question 2: Cross-National Consistency of Gender Effects

As indicated in Table 1, the ANOVAs revealed significant Gender effects for IBQ-R Surgency, ECBQ Surgency, ECBQ Regulatory Capacity, and all CBQ dimensions; and Gender by Nation interactions were significant for all IBQ-R and ECBQ dimensions. Consistent with previous studies (e.g., Else-Quest et al., 2006), males were rated higher in Surgency, and females were rated higher on Negative Affectivity and Regulatory Capacity. Effect sizes for gender were larger for the CBQ than the infant and toddler measures, and more pronounced for Regulatory Capacity than other dimensions.

The nature of the significant interactions was probed through tests of simple effects for Gender (i.e., ANOVA with gender as IV and age as covariate) across all nations. The results of these nation-specific tests are available online in supplemental file 3. For IBQ-R Surgency, males were rated significantly higher than females in Canada, Finland, Israel, and Taiwan, but lower in Russia. For IBQ-R Negative Affectivity, females were rated significantly higher in Canada, the Czech Republic, Finland, Malta and the US, and lower in none. For IBQ-R Regulatory Capacity, females were rated significantly higher than males in Belgium, Finland, Germany, Latvia, Mexico, Russia and Spain, but lower in Malaysia.

Although the Gender by Nation interactions were not significant for the ECBQ scales, the nation-specific tests of gender are nonetheless informative in terms of their consistency. For ECBQ Surgency, males were rated higher than females in 5 of the thirty-nine countries and lower in one. For ECBQ Negative Affectivity, females were rated higher than males in 6 of the thirty-nine countries and lower in none. For ECBQ Regulatory Capacity, females were rated higher than males in eighteen of thirty-nine nations and lower in none.

For CBQ scales, the interactions were largely due to differences in magnitude, rather than direction, of effects. For CBQ-Surgency, males were significantly higher in 32 of the 46 countries, and significantly lower in none. For CBQ-Negative Affectivity, females were rated higher in 12 nations and lower only in Colombia. For CBQ-Regulatory Capacity, females were significantly higher in 37 of 46 countries, and lower in none.

Question 3: Relations between Cross-Cultural Patterns Across Infancy and Childhood

The aggregate temperament scores (marginal means) for all nations on the refined IBQ-R, ECBQ and CBQ are contained in Table 2, and used in subsequent analyses. Because scores from the original IBQ-R, ECBQ and CBQ scales are useful for comparisons with other studies relying on these scales, we also provide nation-level aggregate temperament scores for these scores in Table 3.

To investigate the nature of relations between nations' aggregate temperament scores when assessed at different ages, countries' marginal means for all scales at the three age ranges were correlated. As shown in Table 4, consistency was evident across all ages for Negative Affectivity, and from infancy to toddlerhood for Regulatory Capacity. Surgency scores were not related across age/instrument. However, consistent with analyses at the individual level by Putnam et al (2008), high IBQ-R Surgency was marginally linked to high ECBQ Regulatory Capacity; and also positively correlated with CBQ Negative Affectivity.

Question 4: Relations with Personality Findings

Table 5 contains correlations between countries' marginal means omnibus scores (i.e., marginal means for Surgency, Negative Affectivity and Regulatory Capacity averaged across the

IBQ-R, ECBQ and CBQ; see Table 2) and the self- and other-rated aggregate personality values published by Allik et al. (2017) and McCrae et al. (2005). Correlations with the Omnibus Surgency score indicated that countries in which individuals rated themselves high in Extraversion and Conscientiousness, and low in Neuroticism, rated their children high in Surgency. Correlations with Omnibus Negative Affectivity suggest that countries in which individuals rated themselves high in Neuroticism and rated other adults as low on Openness viewed their children as high in Negative Affectivity. A marginal association also suggested that low self-rated Openness was also associated with high temperamental Negative Affectivity. No correlations were significant between aggregate Regulatory Capacity and personality.

Question 5: Relations with Cultural Orientation and National Wealth

Correlations between omnibus temperament scores and country scores for Hofstede's six cultural orientation dimensions, and for GNI-PC, are shown in Table 6. Because we wished to determine whether different cultural orientations and national wealth predicted unique variance in aggregate temperament scores, results of multiple regression using the six cultural dimensions and GNI-PC as predictors are also presented.

Correlations suggested marginal negative associations between Surgency and both Long/Short-Term Orientation and GNI per capita, with the association between Surgency and Long-Term Orientation becoming significant in the regression analysis. Correlations indicated countries whose children were rated as high in Negative Affectivity had cultural orientations reflecting Collectivism, high Power Distance, Short-Term Orientation and low GNI-PC. Effects for Power Distance and Long/Short-Term Orientation became nonsignificant and marginal, although the effects for Individualism and GNI-PC remained, and a significant positive effect of

Masculinity emerged. Aggregate Regulatory Capacity was not associated with cultural orientation nor GNI-PC.

Question 6: Geographical Patterns of Aggregate Temperament

To facilitate qualitative comparisons of geographical regions, the omnibus scores for each country (see Table 2) are reflected in shading on Figures 1-3. Our interpretations were aided through consideration of marginal means resulting from UN region by sex ANOVAS, with age covaried, of the scale scores from the three questionnaires (see Table 7). Below, we discuss the general trends evident in these three maps, also noting findings that differed by age.

Across questionnaires, Surgency was consistently high across Southeastern Asia, Australia/New Zealand, Northern America and South America; and low in Eastern Asia and Eastern Europe. For other regions, trends were often inconsistent across questionnaire. In particular, Southern Asia demonstrated high Surgency on the CBQ but low Surgency on the IBQ-R; while Central America (Mexico) demonstrated the opposite pattern. Due to the lack of nation-level consistency across age period/measure (described with respect to Question 3 above), nation-level trends are not discussed.

Relatively clear patterns were apparent for Negative Affectivity. On all questionnaires, parents from Southern Asia, South America, and Western Asia and Southern Europe and tended to report high levels of negative affect in their children. In contrast, very low levels of negativity were consistently reported across Northern and Western Europe. At the level of nation, Colombia, Kosovo, Iran, Pakistan, the Philippines, Malaysia, Nigeria, and Türkiye exhibited considerably high Negative Affectivity on the IBQ-R, ECBQ and/or CBQ; while low Negative Affectivity was reported in the Netherlands, Denmark, Sweden, Switzerland, Ireland, Finland and Curacao.

In general, reports of Regulatory Capacity were highest in Southeastern and Western Asia; and low in Eastern Asia for all questionnaires. Notable inconsistencies across questionnaire were observed in Southern Asia and Australia/New Zealand, in which infants and toddlers rated relatively low, but older children were rated high; while the opposite pattern was apparent in Eastern Europe. Regional inconsistency was notable in South America, with very high and very low scores for different nations. Countries showing high omnibus Regulatory Capacity included Malta, Argentina, Bosnia and Herzegovina, Suriname, Denmark, Hungary, Peru and Serbia. Nations rating infants and/or children low in Regulatory capacity included Colombia, Japan, Brazil, and China.

Discussion

The GTP represents a unique collaborative effort to investigate links between broad societal forces and early appearing individual differences. Results of the current report indicate that these links are considerable at all ages tested, but also complex, demonstrating consistent worldwide patterns across early life periods for traits involving negative affect but not those associated with active and approachful behavior. Nation-level differences in temperament were larger than those obtained in similar studies of adult personality, also demonstrating modest consistency with the results of such investigations. Global patterns suggest that reports of high surgency were characteristic of cultures emphasizing short-term goals; while low levels of negative emotionality were particularly common in northern and western Europe, and in wealthy countries that promote individualist values; in contrast to areas of southern Asia and South America. Analyses of the GTP also informed knowledge regarding gender differences in temperament, which grew in consistency and magnitude from infancy through early childhood.

Given the conceptual distance between the historical and philosophical factors shaping national cultures and those organizing individual human development, the magnitude of these relations was surprising. Effect sizes for nation were somewhat larger than those from similar investigations of personality. The scale of these relations among distal phenomena is similar to, or greater than, those reported in more traditional studies of proximal influences. For instance, in a recent meta-analysis by van Dijk et al. (2020) of associations among interparental conflict, parenting, and child adjustment, the average correlation between parent and child variables was .20, corresponding to an effect size nearly half of the average effect of nation in the GTP data. These comparisons suggest that dissimilarities in ways that parents in a given nation enact discipline and offer support to their offspring may have less influence on their children's developing personality than the macrosystem forces that cause most members of their national culture to transmit a common set of values and concerns.

The most powerful and reliable findings connecting culture to temperament were those involving children's Negative Affectivity. The effect of nation on ECBQ and CBQ scores was considerably larger for Negative Affectivity than for Surgency or Regulatory Capacity. Moreover, countries in which infants were viewed as expressing high levels of negative emotions were largely the same as those identifying frequent and strong distress in toddlers and children. The relative consistency and strength of these findings for Negative Affectivity may reflect the salience of emotional displays to parents, as well as nation-level differences in the prevalence of alleles associated with sensitivity to social evaluation (e.g., Way & Lieberman, 2010). The special nature of this aspect of temperament is apparent in the importance it was given in the earliest studies of temperament (e.g., Thomas & Chess, 1977), and although the elicitors of sadness, anger, and fear may change over the early years of life, the appearance of negative emotions is similar in infants and older children, and readily observed by parents. Robust findings for Negative Affectivity may also reflect consistency of the cultural underpinnings governing these tendencies. Societal views of the acceptability of negative emotion displays may guide parental responses that reward or discourage such displays, and/or may contribute to bias in parent's judgments of the frequency, latency, and intensity of negative emotions in their offspring. In contrast, lower effect sizes and poor stability for Surgency, and of Regulatory Capacity between early and late childhood, may reflect changes in how they are measured or shifting expectations for enthusiastic and restrained behavior in several countries in response to dramatic changes in physical, attentional, and self-control capacities over the first decade of life.

Inconsistencies in measurement and developmental cascades, in which one attribute shapes the development of another, may explain unexpected associations between aggregate temperament scores and aggregate self- and other-rated personality. Whereas the findings of high Surgency in countries exhibiting high adult self-rated Extraversion, and relations between Negative Affectivity and self-rated Neuroticism, were expected due to conceptual similarity between these constructs, other temperament-personality associations are more difficult to explain. Low self-rated Neuroticism and high self-rated Conscientiousness also predicted high Surgency, and low other-rated Openness predicted high Negative Affectivity. Country-level associations between low Surgency and high Neuroticism may reflect the relevance of fear (or fearlessness) for engagement in several behaviors assessed in (reversed) Surgency items, leading this scale to be somewhat akin to Big Five Neuroticism, which involves perceptions of the world as threatening. Similarly, societal beliefs influencing high levels of negativity during infancy and childhood may also lead to a lack of comfort with novel experiences in adults, leading to the inverse relation between Negative Affectivity among infants and children and other-rated Openness in adults.

The most basic and fascinating question inspiring the GTP and other studies in this vein is simply “How are people (in this case, infants, toddlers and children) different around the world?” The answers provided by GTP replicate and extend two previous multi-nation investigations of this question (Gartstein & Putnam, 2018; Putnam & Gartstein, 2017). As in the earlier studies, a strong pattern emerged in which low aggregate Negative Affectivity was evident in Northern and Western Europe, while relatively high aggregate Negative Affectivity regions included South America and Southern Asia. The physical distance between these latter regions belies similarities in their cultural orientation, which tend toward Collectivism, high Power Distance and/or Short-Term Orientation.

Our findings regarding high levels of negative affect in infants, toddlers and children from collectivist and power distant nations were consistent with those reported by Putnam and Gartstein (2017), who suggested that caregivers’ anticipatory responding to infants’ needs in collectivist cultures (e.g., Greenfield, Keller, Fuligni & Maynard, 2003) reflected greater acceptance of negative emotion displays. High levels of Power Distance, involving an acceptance of inequalities among members of a society, and Short-Term Orientation, reflecting societal importance of addressing immediate needs, may likewise yield developmental niches that support relatively frequent and intense negative emotions. These interpretations, however, are counter to research indicating that mothers from South Asian countries were likely than those from Western nations to respond with nonsupportive and minimizing reactions in response to their children’s distress. (McCord & Raval, 2016; Trommsdorff, Cole & Heikamp, 2012; also see Raval & Walker, 2019), and to findings indicating that parents in Power Distant cultures such as those in Latin America emphasize the importance of demonstrating respect to family members (see review by Halberstadt & Lozada, 2011). It is a challenge to resolve our findings of higher negative affect in cultures in which such displays are discouraged. Response bias may contribute to this discrepancy. It is plausible that parenting philosophies leading parents to report limiting negative emotion expression in their children also cause them to interpret their offspring's behavior as indicating more frequent and intense negative emotions. Another resolution concerns specific emotions. These cultural prohibitions may be primarily relevant to anger directed at parents, especially with increasing child age, whereas fearfulness, discomfort and sadness in infants and young children may be more accepted, possibly viewed as requests for comfort from more powerful social partners. Indeed, Ravel (Ravel et al., 2016; Ravel & Martini, 2009) found anger to be considered sadness to be more objectionable than other forms of negativity. Whereas the current exercise collapsed Negative Emotionality across several forms of affect, future analyses using GTP samples for which Short- or Standard-Form data with separate scales for these forms of negativity would be useful for indicating more fine-grained associations between cultural orientation and the expression or regulation of emotion. Also, although observational research regarding shyness among Chinese than Canadian children (e.g., Chen et al., 1998) is consistent with our results regarding greater negative affect in non-Western children, extension of observational methods to explore cultural differences in other negative emotions in conjunction with parents' responses, will be valuable in resolving this apparent discrepancy.

Parents from countries in eastern Asia reported very low levels of Surgency. These results are consistent with several previous studies. In early cross-national investigations, Hsu et al. (1981) reported low activity levels and approach tendencies in Taiwan in comparison to US infants, Windle et al. (1988) found lower positive affectivity in preschoolers from Japan in comparison to US children, and Ahadi et al. (1993) indicated lower activity, approach, and high-intensity pleasure in Chinese compared to US children. These strong tendencies for low activity and approach may have roots in Long-Term Orientation. This cultural dimension was initially developed through the Chinese Value Survey, and the values of thrift, perseverance and a sense of shame were interpreted as reflecting values inherent to Confucianism. Behaviors associated with Surgency, including strong desires for anticipated rewards and seeking of intense experiences, may be counter to such values. Indeed, Bond and Wang (1983, as cited in Hofstede et al., 2010) suggested that self-assertion and expectation of immediate gratification were to be discouraged among children in these societies. The current findings suggest that practices and perceptions of parents from countries emphasizing long-term goals are reflected in the less active and exuberant conduct of their young children.

The research described above connecting parenting to cultural orientation, similar to most cross-cultural temperament literature, is inconclusive, as it has tended to rely on comparisons between few countries that differ on multiple aspects of culture. To our knowledge, only one study has explicitly examined parenting practices in relation to Hofstede’s dimensions. In the context of the JETTC, Gartstein and Putnam (2019) found parents in individualist and low power distant nations endorsed gentle sleep encouragement methods and decreased emphasis on guilt-inducing discipline techniques. More extensive cross-cultural research with samples reminiscent of the GTP would enhance understanding of how national values might be translated through parents’ beliefs and actions to shape their children’s emotional and social proclivities.

High levels of Negative Affectivity was additionally associated with low national wealth. Hofstede and McCrae (2004) also found relations between GNP and aggregate personality, although the direction of these findings is somewhat contrary to the current study, as national wealth was positively correlated with Extraversion, unrelated to Neuroticism and negatively correlated with Conscientiousness. More consistent with our findings are results of Strickhouser and Sutin (2019), in which lower family SES and neighborhood SES independently predicted lower sociability, higher reactivity and lower persistence in a sample of Australian children. Strickhouser and Sutin (2019) suggested that the chronic stress and anxiety that often accompany poverty may influence the development of emotional and motivational systems underpinning temperament. These environmental variables may influence the development of emotional and motivational systems underpinning temperament through effects on the development of neural circuitry involving the amygdala, frontal cortex and HPA axis (Assari, Boyce & Bazargan, 2020; Noble & Giebler, 2020). Analyses of the GTP herein suggest that, even at a national level, it may be adaptive for children at lower income levels to more frequently experience and demonstrate negative emotions than their agemates in more affluent nations. Extending the approach of Strickhouser and Sutin (2019) to a multinational perspective would yield insight regarding the most critical aspects of environmental stress for emotional development.

Curiously, although nation-level differences explained substantial variance in Regulatory Capacity at all ages, countries demonstrating high or low levels of regulation at one age were only consistent from infancy through toddlerhood, and this aspect of temperament was unrelated to adult personality, cultural orientation and national income. The lack of association between regulation and individualism-collectivism is particularly surprising given emphasis on autonomy-promotion in individualist cultures in comparison to greater levels of control among parents in collectivist cultures (see Rothbaum & Trommsdorff, 2007). Empirical associations between parenting, self-control and individualism, however, are inconsistent. For instance, Chen-

Bouck, Patterson and Chen (2019) found collectivism to be positively correlated with some aspects of parental control, but negatively correlated with others; and Li, Vazsonyi and Dou (2018) found lower attitudinal self-control, but higher behavioral self-control in Chinese, compared to US college students. Complex links between different forms of regulation and cultural orientation warrant further exploration.

Although development of gender differences in temperament was not a primary focus of the GTP, our analyses were nonetheless informative. In general, our findings confirmed findings from prior large-scale efforts (e.g., Else-Quest et al., 2006) showing boys to be higher in surgency, with girls higher in regulatory capacity and slightly higher in negative affectivity. These differences were modest or nonsignificant in infancy. At later ages, differences were far more pronounced. Although these differences varied by nation, this nation-level variability accounted for only a very small portion of variance. This relative consistency in direction across countries may reflect common biological forces playing out in youth around the world and/or a degree of worldwide consistency in parental expectations and bias in ratings. Increases in magnitude over childhood, in contrast, suggest the compounding power of socialization as children increasingly perceive and actualize expectations regarding gender roles, or may reflect maturation of biological systems organizing sex differences.

The results of the GTP provide confirmation of the surprising relations between the reactivity and regulation demonstrated in infants and young children and the geographical and cultural contexts that organize their developmental niches. Like all correlational results, however, they only invite speculation regarding the origin and causal direction of these differences. In our writing, we have largely emphasized a framework common to socialization science, in which environments shape attributes of individuals in a given locale, or at least the way in which parents report on the behavior of their offspring. Other causal pathways, however, are plausible. McCrae (in Hofstede & McCrae, 2004) suggests two possibilities: selective migration, in which individuals, groups and families move to regions in which their characteristics are more valued; or reverse causation, in which the traits frequently demonstrated by members of a society, including those that appear early in life and are genetically influenced, guide the creation of laws, habits and values adopted by their culture. Consistent with these suggestions, the relative proportion in different nations of polymorphisms of genes including 5-HTTLPR, A118G and MAOA-UVNTR, conceived as modulating individuals' sensitivity to social evaluation, have been linked to cultural dimensions including Individualism/Collectivism and Long-Term/Short-Term Orientation (Chaio & Blizinsky, 2010; Minkov et al., 2015; Way & Lieberman, 2010). Both directions are presumably relevant and are coordinated in the "gene-culture coevolutionary theory", in which cultural values are viewed as having evolved, both reflecting and influencing the environments (social and physical) under which genetic selection takes place in a dynamic and reciprocal manner (Boyd & Richerson, 1985; Chiao & Blizinsky, 2010; Putnam & Gartstein, 2018; Rentfrow et al., 2008). We view temperament and personality as an intermediate process and suggest a "gene-trait-culture" model, in which enculturation and selection pressures shape characteristic ways of acting and thinking, forming cultural orientation processes that feed back into socialization, migration, and reproduction patterns shaping aggregate characteristics of populations.

Enthusiasm for these surprising connections between macrosystem dimensions and individual differences must be tempered by consideration of their limitations. Although the effect sizes for nation-level culture on temperament are similar to or larger than those found for more proximal variables, less than 10% of the variance was explained for many of the temperament variables. As such, the variability of individuals within national cultures is far greater than differences among the average members of societies. The greatest danger inherent to results involving group comparison is the potential for misapplication and stereotyping, and readers are exhorted to acknowledge the exquisite diversity of persons living within any culture. Potential for misapplication of findings is enhanced by tendencies to place value statements on temperament traits. Because a frequent use of temperament scores has involved prediction of outcomes such as adjustment and academic performance, certain temperament attributes are sometimes considered as nonoptimal risk factors. Although these temperament-adjustment connections are roughly consistent across and between multiple cultures (González-Salinas et al., 2018), this limited view does not lend “good” or “bad” assignations to national-level data. Rather, attributes frequently demonstrated in different locales merely indicate common adaptations of individuals to their home context, and should be interpreted as informative, not pejorative.

An apparent concern regarding these findings is the temporal gap between the measurement of adult personality and cultural orientation in comparison to temperament. Rather than a weakness, we view this aspect of our findings as an intriguing phenomenon. The fact that cultural dimensions collected from discrete slices of national populations (IBM employees) are associated with the perceived behavior of these nations’ youngest citizens 50 years later speaks to the importance of these shared values. This contrasts with recent research and common observations that speak to the notion of cultural discontinuity. Increased globalization is changing the contexts in which all people on earth develop, with change occurring in different ways in different communities. These changes have ramifications for temperament and parenting. A notable example is research demonstrating change in the correlates of temperament in Chinese populations. Research in the 1990s (e.g., Chen et al., 1992) found shy and sensitive behavior in children to be associated with compromised peer relationships in Canadian children, but with leadership and peer acceptance for Chinese youth. However, these relations changed over time in China: in comparison to findings from the sample assessed in 1990, a sample assessed in 2002 found shyness to be associated with peer rejection, school problems and depression (Chen et al., 2005). Ongoing research suggests these trends have continued and extend to rural areas of China, with comparisons of 2012 and 2022 cohorts showing increasingly negative relations between shyness and social and academic competence (Li et al., 2023). Parenting practices have similarly changed, with Chinese parents expressing greater encouragement of child autonomy and becoming less involved in children's activities between 1995 and 2008 (Chen et al., 2021). Hofstede (Hofstede et al., 2010) has proposed cultural change as occurring through a series of layers, with observable practices (perhaps including parenting and peer impressions) that can be altered within an individual's lifetime comprising an outer, more flexible layer, while values of the type represented by cultural orientation are transferred more fully from generation to generation. Consistent with this notion, relative differences between nations have remained stable over several decades (Beugelsdijk et al., 2015). It is possible that nation-level temperament may be implicated in the stability of these cultural characteristics. Extending scholarship of the type represented in this report over coming decades may reveal the impact of historical time on cultural orientation, temperament, parenting, and their associations.

The origins of geographical distribution of temperament differences remain obscure. Throughout this discussion, we have suggested that aspects of culture captured in Hofstede's dimensions may have meaningfully influenced by, and are reflected in, the perceived behavior of infants and children. This interpretation is called to question by Galton's problem, a form of autocorrelation in which causally unrelated constructs covary with one another due to their manifestation in geographically close areas. As such, similar temperament and cultural orientation in geographically proximate areas may have resulted from historical connections among the peoples populating these subregions. We have attempted to address Galton's problem by interpreting geographical trends in temperament according to broad global regions in addition to analyses of nation-level scores. This approach suggested that relations between temperament and cultural orientation are not completely explained by geography. For instance, parents from collectivist nations in Asia and South America both observed high levels of negative affectivity in their offspring, suggesting a more substantially meaningful relationship between collectivism and emotion. Our use of region is consistent with previous approaches that restrict analyses to a limited number of geographically disparate cultures or regions (e.g., Korotayev & de Munck, 2003; Ross & Homer, 1976). More recent approaches (e.g., Zhang, Lee, DeBruine & Jones, 2019) utilize MLM to nest nations within region, a technique we were unable to exploit due to limited power. Even within these approaches, a central concern inherent to Galton's problem remains, in that the mechanism explaining relations between cultural and individual tendencies connecting (for example) collectivism to negative affectivity can remain unclear. As described above, societies both shape and are shaped by the individuals within them, and both individual and cultural tendencies may be shaped by other social and ecological factors. Denton (2007) indicates the importance of clearly specified theoretical models involving multiple predictors to disentangle these complex relations. We strongly encourage future efforts guided by explicit consideration of historical, linguistic, economic and other factors that impact both individuals and societies.

The strongest caveat when considering these findings concerns reliance on parent report. Throughout this manuscript, we have attempted to interpret patterns as reflecting either actual child behavior or response bias. Contemporary views on temperament measurement (e.g., Rothbart & Bates, 2006) recommend a “components-of-variance” approach, acknowledging limits of questionnaire data while maintaining that such measures also contain valid information about actual behavior. A measurement concern more unique to cross-cultural investigation is the reference group effect, in which individuals are not rated against a universal norm, but in comparison to other members of their culture (Heine et al., 2002). The results of previous cross-cultural studies using observational data (e.g., Chen et al., 1998) converge modestly with questionnaire findings. Valuable tasks for future studies are to identify and utilize observational measures appropriate for multiple cultures. Scores resulting from such measures could be productively used to identify discrepancies between cultural effects on objective ratings and subjective perceptions of parents and others to better understand relations between what adults in different cultures desire and expect from their children regarding displays of emotion and how their children actually behave. An alternative, complementary approach is to explicitly measure cultural expectations about normative child behavior, in a manner similar to Terracciano et al. (2005). Studies of this nature can also be complemented by collection of genetic polymorphisms and peripheral biobehavioral markers (e.g., vagal tone) associated with temperamental differences.

The current study, although broader in scope than prior efforts, is nonetheless limited by the number and nature of nations included. Only 59 of the 195 countries recognized by the United Nations contributed data to the GTP. A glance at Figure 1 reveals the inadequate inclusion of reports from central Asia, the Middle East and nearly all of Africa. Greater outreach to potential collaborators in these regions will improve future iterations of the GTP and related datasets.

Furthermore, the degree to which the 341 datasets used are truly representative of the 59 nations they represent is uncertain. Over one-third of the GTP countries are represented by a single dataset, and in all countries, the majority of the datasets acquired are from convenience samples, often measured in relative proximity to a major university, and are likely to be more well-educated, wealthy, and racially/ethnically homogenous than the larger populations from which they were drawn. Still, several GTP nations are represented by samples from geographically distinct areas (see Supplemental Materials). Research demonstrating substantial within-nation differences between suburban and rural/traditional parents in emotion socialization practices and philosophies (e.g., Raval & Martini, 2009) suggests divergent developmental niches that may promote different temperament traits. Adult personality has been shown to vary meaningfully within a given country (e.g., Rentfrow et al., 2008), and, as indicated in the methods section, at least 14 generated temperament scores in the GTP differed substantially from the national aggregates. More detailed within-nation investigations of differences between samples from the GTP provide an intriguing future direction that may isolate factors contributing to such variability while informing understanding of the current analyses.

Concerns also exist regarding the reliance on the three factors commonly derived from the IBQ-R, ECBQ, and CBQ. Although multiple studies from various cultures have utilized these factors, variations have also been identified between nations (e.g., Ahadi et al., 1993; Gartstein et al., 2005). In the current study, confirmatory factor analyses indicated poor fit of the three-factor model of the ECBQ and CBQ, and fit diminished for all three measures when the configural model was tested across 20 large samples. This failure to fit simple structure is unsurprising, given the complexity of temperament displays and the development of these measures. The psychobiological scales were developed using the rational method, with items generated regarding behaviors believed to reflect fine-grained dimensions, but no effort was made to avoid items that may reflect other dimensions. In addition, the 3-factor model was not hypothesized, but instead emerged in an inductive fashion (Rothbart et al., 2001). These steps led to items sharing error variance with other items from outside their initial scale or factor in the very short forms. For instance, modification indices from our analyses suggested fit of the CBQ could be improved by allowing the error term for item 30 (Approaches places cautiously; on the REG factor, originally from the Inhibitory Control scale) to correlate with the error for item 31 (Slow and unhurried when deciding what to do next; on SUR, originally from Impulsivity). It is likely that the behaviors measured with these items involve both surgent, impulsive behavior balanced against regulatory restraint. Although the three-factor model has been useful in organizing understanding temperament structure, and has underlied the organization of hundreds of empirical articles, more thorough refinement of the measures may yield independent dimensions that can be assessed across diverse samples.

Regardless of the remaining questions concerning temperament structure, we were able to derive scales that measured similar latent factors across the several nations represented in the GTP. Our investigation of measurement invariance, however, was compromised by the large scale of the project. Tests of measurement invariance are typically carried out through comparisons of only two or three samples. Byrne and van de Vijver (2010) provided a model for testing weaker forms of invariance (i.e., configural and metric) across multiple samples, which we followed in our analyses, but characterized their own procedures as “impractical under normal circumstances” (p. 128) and did not attempt to develop methods to assess stronger forms of invariance across multiple groups. Because we were unable to explore scalar or error invariance that would indicate the degree to which differences in latent scores underlying scales are uniformly manifest in all items, it is possible that these broad scales gloss over important distinctions within their parameters. One likely source of such variance are the fine-grained constructs that comprise Surgency, Negative Affectivity and Regulatory Capacity. Methodological advances that facilitate more stringent tests between multiple samples, and more detailed studies involving the fine-grained scales comprising the standard and short forms of the instruments in relation to cultural orientation and other variables represent important future directions to both confirm and elaborate upon the findings reported herein.

The relatively large number of countries and communities making up the GTP also hold promise for future investigations of physical and social contributors to temperament development. Again, existing adult personality datasets provide a valuable model. The country-level means reported by McCrae (2002) have been examined in relation to a variety of constructs, ranging from cancer rates and substance abuse to corruption and pathogen spread (Connelly & Ones, 2008; McCrae & Terracciano, 2008; Schaller & Murray, 2008). Exploration of temperament in relation to latitude, wealth, religion, cultural orientation, and other factors may foreshadow and guide studies regarding influences and outcomes linked to global patterns of temperament.

A final direction for future studies is consideration of potential mediators and moderators. For instance, preliminary findings from Gartstein and Putnam (2019) suggest that relations between national scores on collectivism and behavior problems are statistically reduced by inclusion of variables reflecting techniques used by parents to put their children to sleep; and Lansford et al. (2005) found that the impact of physical discipline on child adjustment differs according to the degree it is perceived as normative in a culture. Cultural factors may also moderate biological mechanisms. For instance, whereas previous research with Western samples (e.g., Arbelle et al., 2003) had indicated higher shyness among children with the “long” allele of 5-HTTPLR, Chen et al. (2014) found the opposite pattern in a sample of Chinese children. The large number of collaborators and communities represented by the GTP constitutes a valuable resource for more studies of this type.

The incredible variability of human behavior is vividly portrayed in the current study. The national cultures included in this effort differ dramatically in their values and goals. The young individuals whose ratings make up these data differ in relation to these cultural leanings, yet the large majority of variance among them is caused by biological and environmental factors not assessed here. In contrast to this focus on variability, the most important lesson to emerge from this effort is one of unity. The graciousness of the hundreds of researchers who joined in this endeavor by including their data is humbling and inspiring. Beyond the opportunities presented by their collective body of data, the GTP holds promise for continued partnerships that enable a deeper understanding of how we all come to be the persons we become.

References

Achenbach, T. M., & Rescorla, L. A. (2000). *Manual for the ASEBA Preschool Forms & Pres*. Burlington, VT: University of Vermont Department of Psychiatry.

Ahadi, S.A., Rothbart, M. K., & Ye, R. (1993). Children's temperament in the US and China: Similarities and differences. *European Journal of Personality, 7*, 359-378. doi:10.1002/per.2410070506

Allik, J., Church, A.T., Ortiz, F. A., Rossier, J., Hrebickova, M., de Fruyt, F., Realo, A. & McCrae, R. R. (2017). Mean profiles of the NEO personality inventory. *Journal of Cross-Cultural Psychology, 48*, 402-420. doi:10.1177/0022022117692100

Allik, J. & McCrae, R. R. (2004). Towards a geography of personality traits: Patterns of profiles across 36 cultures. *Journal of Cross-Cultural Psychology*, *35*, 13-28. doi:10.1177/0022022103260382

Allport, G. W. (1961). *Pattern and growth in personality*. New York: Holt, Rinehart & Winston.

Arbelle, S., Benjamin, J., Golin, M. Kremer, I., Belmaker, R. H., & Ebstein (2003). Relation of shyness in grade school children to the genotype for the long form of the serotonin transporter promoter region polymorphism. *American Journal of Psychiatry, 160*, 671-676. doi: 10.1176/appi.ajp.160.4.671.

Barcenilla, C. C., Luttges, B. L., Rojas-Barahona, C. A., & Campos, A. L. (2021). Psychometric analysis of the Children's Behavior Questionnaire (CBQ) in Chile. *Current Psychology: A Journal for Diverse Perspectives on Diverse Psychological Issues*, May 25, 2021. doi:10.1007/s12144-021-01871-9.

Bartram, D. (2013). Scalar equivalence of OPQ32: Big Five profiles of 31 countries. *Journal of Cross-Cultural Psychology, 44*, 61-83. doi:10.1177/0022022111430258

Bates, J. E., Freeland, C. A. B., & Lounsbury, M. L. (1979). Measurement of infant difficultness. *Child Development*, *50*, 794–803. doi:10.2307/1128946

Beugelsdijik, S., Maseland, R. & van Hoorn (2015). Are scores on Hofstede's dimensions of national culture stable over time? A cohort analysis. *Global Strategy Journal, 5*, 223-240.

Bond, M. H. & Wang, S. H. (1983), Aggressive behavior and the problem of maintaining order and harmony. In A. P. Goldstein & M. H. Segall (eds.) *Global Perspectives on Aggression* (pp. 58-73). New York : Pergamon.

Boyd, R., & Richerson, P. J. (1985). *Culture and the evolutionary process*. Chicago: University of Chicago Press.

Bronfenbrenner, U. (1979). *The Ecology of Human Development*. Cambridge: Harvard University Press.

Burman, J. (2019). Development. In R. J. Sternberg & W. E. Pickren (Eds) *The Cambridge Handbook of the Intellectual History of Psychology* (pp. 287-317). Cambridge: Cambridge University Press. doi:10.1017/9781108290876.012.

Byrne, B. M., & van De Vijver, F. J. R. (2010). Testing for measurement and structural equivalence in large-scale cross-cultural studies: Addressing the issue of nonequivalence. *International Journal of Testing*, *10*(2), 107–132. doi: 10.1080/15305051003637306

Chen, X. (2000). Growing up in a collectivist culture: Socialization and socioemotional development in Chinese children. In A. L. Comunian & U. P. Gielen (Eds.), *International perspectives on human development* (pp. 331–353). Pabst Science Publishers.

Chen, X. (2018). Culture, temperament, and social and psychological adjustment. *Developmental Review*, *50*, 42-53. doi:10.1016/j.dr.2018.03.004

Chen, X., Cen, G., Li, D. & He, Y. (2005). Social functioning and adjustment in Chinese children: The imprint of historical time. *Child Development*, *76*, 182-195.

Chen, X., Chen, X., Zhao, S., Way, N., Yoshikawa, H.,…Li, D. (2021). Autonomy- and connectedness-oriented behaviors of toddlers and mothers at different historical times in urban China. *Developmental Psychology, 57*, 1254-1260.

Chen, X & French, D. C. (2008). Children's social competence in cultural context. *Annual Review of Psychology, 59*, 591-616. doi:10.1146/annurev.psych.59.103006.093606

Chen, X., Zhang, G., Liang, Z, Zhang, M., Way, N., Yoshikawa, H., Xiaoyan, K., Ke, X., Lu, Z., & Deng, H. (2014). The association between 5-HTTLPR gene polymorphism and behavioral inhibition in Chinese toddlers. *Developmental Psychobiology, 56*, 1601-1608.

Chen, X., Hastings, P. D., Rubin, K. H., Chen, H., Cen, G., & Stewart, S. L. (1998). Child-rearing attitudes and behavioral inhibition in Chinese and Canadian toddlers: A cross-cultural study. *Developmental Psychology, 34*, 677. doi:10.1037/0012-1649.34.4.677

Chen, X., Liu, M. & Qinglin, B. (2022). Culture and Children's Social Development. In P.K. Smith & C.H. Hart (Eds). *The Wiley-Blackwell Handbook of Childhood Social Development* (pp. 241-259. Hoboken, NJ: Wiley Blackwell.

Chen, X., Rubin, K. H., Liu, M., Chen, H., Wang, L., Li, D. ..., Li, B. (2003). Compliance in Chinese and Canadian toddlers: A cross-cultural study. *International Journal of Behavioral Development, 27*, 428-436. doi:10.1080/01650250344000046

Chen, X., Rubin, K. H. & Sun, Y. (1992). Social reputation and peer relationships in Chinese and Canadian children: A cross-cultural study. *Child Development, 63*, 1336-1343.

Chen-Bouck, L., Patterson, M. M. & Chen, J. (2019). Relations of collectivism socialization goals and training beliefs to Chinese parenting. *Journal of Cross-Cultural Psychology, 50*, 396-418.

Chiao, J. Y. & Blizinsky, K. D. (2010). Culture-gene coevolution of individualism-collectivism and the serotonin transporter gene (5-HTTLPR). *Proceedings of the Royal Society B: Biological Sciences, 277*, 529-537. doi:10.1098/rspb.2009.1650

Connelly, B. S. & Ones, D. S. (2008). The personality of corruption: A national-level analysis. *Cross-Cultural Research, 42*, 353-385. doi:10.1177/1069397108321904

Costa, R. & Figueiredo, B. (2018). Infant Behaviour Questionnaire – revised version: A psychometric study in a Portuguese sample. *Journal of Reproductive and Infant Psychology, 36*, 207-218. doi:10.1080/02646838.2018.1436752

Costa, P. T. Jr., Terracciano, A., & McCrae, R. R. (2001). Gender differences in personality traits across cultures: Robust and surprising findings. *Journal of Personality and Social Psychology, 81*, 322-331. doi:10.1037/0022-3514.81.2.322

Cozzi, P., Putnam, S. P., Menesini, E., Gartstein, M. A., Aureli, T., Calussi, P., & Montirosso, R. (2013). Studying cross-cultural differences in temperament in toddlerhood: United States of America (US) and Italy. *Infant Behavior and Development, 36*, 480–483. doi:10.1016/j.infbeh.2013.03.014

De Bolle, M., De Fruyt, F., McCrae, R. R., L€ockenhoff, C. E., Costa, P. T., Aguilar-Vafaie, M. E.…Terracciano, A. (2015). The emergence of sex differences in personality traits in early adolescence: A cross-sectional, cross-cultural study. *Journal of Personality and Social Psychology, 108*, 171–185. doi:10.1037/a0038497

Denton, T. (2007). Yet another solution to Galton’s problem. *Cross-Cultural Research, 41*, 12-45.

Derryberry, D., & Rothbart, M. K. (1988). Arousal, affect, and attention as components of temperament. *Journal of Personality and Social Psychology, 55*, 958-966. doi:10.1037/0022-3514.55.6.958

Ellis, L. K., & Rothbart, M. K. (2001, April). Revision of the Early Adolescent Temperament Questionnaire. Poster presented at the 2001 Biennial Meeting of the Society for Research in Child Development, Minneapolis, Minnesota.

Else-Quest, N. M., Hyde, J. S., Goldsmith, H. H., & Van Hulle, C. A. (2006). Gender differences in temperament: A meta-analysis. *Psychological Bulletin*, *132*, 33-72. doi:10.1037/0033-2909.132.1.33

Evans, D. E., & Rothbart, M. K. (2007). Developing a model for adult temperament. *Journal of Research in Personality, 41*, 868-888. doi:10.1016/j.jrp.2006.11.002

Gaias, L. M., Raikkonen, K., Komsi, N., Gartstein, M. A., Fisher, P. A., & Putnam, S. P. (2012). Cross-cultural temperamental differences in infants, children, and adults in the United States of America and Finland. *Scandinavian Journal of Psychology, 53*, 119-128. doi:10.1111/j.1467-9450.2012.00937.x

Galindo, C., & Fuller, B. (2010). The social competence of Latino kindergartners and growth in mathematical understanding. *Developmental Psychology, 46*, 579-592. doi:10.1037/a0017821

Gartstein, M. A., Knyazev, G. G., & Slobodskaya, H. R. (2005). Cross-cultural differences in the structure of infant temperament: United States of America (US) and Russia. *Infant Behavior and Development*, *28*, 54-61. doi:10.1016/j.infbeh.2004.09.003

Gartstein., M. A. & Putnam, S. P. (Eds.) (2018). *Toddlers, Parents, and Culture: Findings from the Joint Effort Toddler Temperament Consortium*. London: Routledge Publishing.

Gartstein, M. A., & Rothbart, M. K. (2003). Studying infant temperament via the Revised Infant Behavior Questionnaire. *Journal of Infant Behavior and Development*, *26*, 64-86. doi:10.1016/S0163-6383(02)00169-8

Gartstein, M. A., Slobodskaya, H. R., Żylicz, P. O., Gosztyla, D. & Nakagawa, N. (2010). A cross-cultural evaluation of temperament development: Japan, United States of America, Poland and Russia. *International Journal of Psychology and Psychological Therapy*, *10*, 55-75. https://www.ijpsy.com/volumen10/num1/250/a-cross-cultural-evaluation-of-temperament-EN.pdf

Golmohammadi, G., Zarifian, T., Qhanbari, S., Bakhshi, E. & Sakhai, F. (2020). Cross-cultural adaptation and validation of the Persian version of Children's Behavior Questionnaire in Iranian children. *Current Psychology: A Journal for Diverse Perspectives on Diverse Psychological Issues*, Jul 13, 2020. doi.org/10.1007/s12144-020-00918-7

González-Salinas, C., Sánchez-Pérez, N., Fuentes, L. J., Montirosso, R, & Heinonen, K. (2018). Cross-cultural differences in associations between temperament and behavior problems. In M. A. Gartstein & S. P. Putnam (Eds.) *Toddlers, Parents, and Culture: Findings from the Joint Effort Toddler Temperament Consortium* (pp. 46-54). London: Routledge Publishing. doi:10.4324/9781315203713-5

Gottlieb, G. (1991). Experiential canalization of behavioral development: Theory. *Developmental Psychology, 27*, 4-13. doi:10.1037/0012-1649.27.1.4

Greenfield, P. M., Keller, H., Fuligni, A., & Maynard, A. (2003). Cultural pathways through universal development. *Annual Review of Psychology, 54*, 461–490.

Gudiño, O. G., Lau, A. S. (2010). Parental cultural orientation, shyness, and anxiety in Hispanic children: An exploratory study. *Journal of Applied Developmental Psychology, 31*, 202-210.

doi:10.1016/j.appdev.2009.12.003

Halberstadt, A. G., & Lozada, F. T. (2011). Emotion development in infancy through the lens of culture. *Emotion Review, 3(2)*, 158–168. https://doi.org/10.1177/1754073910387946

Harkness, S., & Super, C. M. (1994). The "developmental niche": A theoretical framework for analyzing the household production of health. *Social Science and Medicine, 38(2)*, 217-226. doi:10.1016/0277-9536(94)90391-3

Heine, S. J., Lehman, D. R., Peng, K., & Greenholtz, J. (2002). What's wrong with cross-cultural comparisons of subjective Likert scales: The reference-group problem. *Journal of Personality and Social Psychology, 82*, 903-918. doi:10.1037/0022-3514.82.6.903

Hofstede, G. (1984). *Culture's Consequences: International Differences in Work-Related Values.* Abridged Ed. Beverly Hills, CA: Sage.

Hofstede, G. (2001). *Culture's Consequences: Comparing Values, Behaviors, Institutions and Organizations Across Nations*. Thousand Oaks, CA: Sage.

Hofstede, G. (2011). Dimensionalizing cultures: The Hofstede model in context. *Online Readings in Psychology and Culture, 2*.doi:10.9707/2307-0919.1014.

Hofstede, G. & Bond, M. H. "The Confucius connection: From cultural roots to economic growth." *Organizational Dynamics 16*, 4-21,

Hofstede, G., Hofstede, G. J. & Minkov, M. (2010). *Cultures and Organizations: Software of the Mind* (Rev. 3rd ed.). New York: McGraw-Hill.

Hofstede, G. & McCrae, R. (2004). Personality and culture revisited: Linking traits and dimensions of culture. *Cross-Cultural Research, 38,* 52-88. doi:10.1177/1069397103259443

Hox, J., & McNeish, D. (2020). Small Samples in Multilevel Modeling. In R. Van De Shoot & M. Miocevic (Eds). *Small Sample Size Solutions* (pp. 215-225). Taylor & Francis.

Kagan, J. (1994). *Galen's prophecy: Temperament in human nature*. New York: Basic Books.

Kaiser, T. (2019). Nature and evoked culture: Sex differences in personality are uniquely correlated with ecological stress. *Personality and Individual Differences, 148*, 67–72. doi:10.1016/j.paid.2019.05.011

Kajonius, P. J., & Johnson, J. (2018). Sex differences in 30 facets of the five factor model of personality in the large public (N= 320,128). *Personality and Individual Differences, 129,* 126–130. doi:10.1016/j.paid.2018.03.026

Keller, H., Lamm, B., Abels, M., Yovsi, R., Borke, J., Jensen, H., . . . Chaudhary, N. (2006). Cultural models, socialization goals, and parenting ethnotheories: A multicultural analysis. *Journal of Cross-Cultural Psychology, 37*, 155–172. doi:10.1177/0022022105284494

Kim, J., Rapee, M. R., Oh, J. K., Moon, H. S. (2008). Retrospective report of social withdrawal during adolescence and current maladjustment in young adulthood: Cross-cultural comparisons between Australian and South Korean students. *Journal of Adolescence, 31,* 543-563. doi:10.1016/j.adolescence.2007.10.011

Kohnstamm, G. A. (1989). Temperament in childhood: Cross-cultural and sex differences. In G. A. Kohnstamm, J. E. Bates and M. K. Rothbart (Eds.). Temperament in Childhood (pp. 483-508. West Sussex, England: John Wiley & Sons.

Korotayev, A. & de Munck, V. (2003). “Galton’s asset” and “Flower’s problem”: Culutral networks and cultural units in cross-cultural research (or, male genital mutilations and polygyny in cross-cultural perspective). *American Anthropologist, 105,* 353-358.

Kotelnikova, Y., Olino, T. M., Klein, D. N., Kryski, K. R. & Hayden, E. P. (2018). Higher- and lower-order factor analyses of the Children’s Behavior Questionnaire in early and middle childhood. *Psychological Assessment, 28*, 92-108. doi:10.1037/pas0000153

Krassner, A., Gartstein, M. A., Park, C., Dragan, W. L., Lecannelier, F., & Putnam, S. P. (2017). East-West, Collectivist-Individualist: A Cross-Cultural Examination of Temperament in Toddlers from Chile, Poland, South Korea, and the U.S. *European Journal of Developmental Psychology, 14*, 449-464. doi:10.1080/17405629.2016.1236722

Lansford, J. E., Chang, L., Dodge, K. A., Malone, P. S., Oburu, P., Palmerus, K., Bacchini, D., Paastorelli, C., Bombi, A.S., Zelli, A., Tapanya, S., Chaudhary, N., Deater-Deckard, K., Manky, B., & Quinn, N. (2005). Physical discipline and children's adjustment: Cultural normativeness as a moderator. *Child Development, 76*, 1234-1246. doi:10.1111/j.1467-8624.2005.00847.x

Lee, E.H., Zhou, Q., Eisenberg, N., & Wang, Y. (2013). Bidirectional relations between temperament and parenting styles in Chinese children. *International Journal of Behavioral Development, 37*, 57-67. doi:10.1177/0165025412460795

Leerkes E. M, Su J., Reboussin, B. A., Daniel, S. S., Payne, C. C., & Grzywacz J. G. (2017). Establishing the measurement invariance of the very Short Form of the Infant Behavior Questionnaire Revised for mothers who vary on race and poverty status. *Journal of Personality Assessment, 99*, 94-103. doi: 10.1080/00223891.2016.1185612.

Li, D., Liu, J., Hui, J., Wu, M & Chen, X. (2023, October). Shyness and social, school, and psychological adjustment in rural Chinese children at different historical times (Conference presentation abstract). Presented at The Occasional Temperament Conference, Orford, Québec, Canada.

Li, J-B, Vazsonyi, A. T. & Dou, K. (2018). Is individualism-collectivism associated with self-control? Evidence from Chinese and U.S. samples. *PLoS ONE 13*: e0208541. https://doi.org/10.1371/journal.pone.0208541

Martin, R.P., Wisenbaker, J. & Huttunen, M. (1994). Review of factor analytic studies of temperament measures based on the Thomas-Chess structural model: Implications for the Big Five. In C.F. Halverson Jr., G. A. Kohnstamm, & R. P. Martin (Eds.), *The developing structure of temperament and personality from infancy to adulthood* (pp. 157-172). Hillsdale, NJ: Erlbaum.

McCord, B. L. & Raval. V. V. (2016). Asian Indian immigrant and White American maternal emotion socialization and child socio-emotional functioning. *Journal of Child and Family Studies*, *25*, 464-474.

McCrae, R. R. (2001). Trait psychology and culture: Exploring intercultural comparisons (2001). *Journal of Personality, 69,* 819-846. doi:10.1111/1467-6494.696166

McCrae, R. R. (2002). NEO-PI-R data from 36 cultures: Further intercultural comparisons. In R. R. McCrae & J. Allik (Eds.). *The five-factor model of personality across cultures* (pp. 105-126). New York: Kluwer Academic/Plenum. doi:10.1007/978-1-4615-0763-5_6

McCrae, R. R., & Terracciano, A. (2008). The five-factor model and its correlates in individuals and cultures. In F. J. R. van de Vijver, D. A. van Hemert, & Y. H. Poortinga (Eds.), *Multilevel analysis of individuals and cultures* (pp. 249–283). Taylor & Francis Group/Lawrence Erlbaum Associates.

McCrae, R.R., Terracciano, A. & 79 Members of the Personality Profiles of Cultures Project (2005). Personality profiles of cultures: Aggregate personality traits. *Journal of Personality and Social Psychology, 89*, 407-425. doi:10.1037/0022-3514.89.3.407.

McCrae, R. R., Terracciano, A., & 78 Members of the Personality Profiles of Cultures Project. (2005). Universal Features of Personality Traits From the Observer's Perspective: Data From 50 Cultures. *Journal of Personality and Social Psychology, 88(3)*, 547–561.

https://doi.org/10.1037/0022-3514.88.3.547

McCrae, R. R., Terracciano, A., De Fruyt, F., De Bolle, M., Gelfand, M. J., Costa, P. T., Jr., & 42 Collaborators of the Adolescent Personality Profiles of Cultures Project. (2010). The validity and structure of culture-level personality scores: Data from ratings of young adolescents. *Journal of Personality,78*, 815–838. doi:10.1111/j.1467-6494.2010.00634.x

McCrae, R. R., Terracciano, A., Realo, A. & Allik, J. (2007). Climatic warmth and national wealth: Some culture-level determinants of national character stereotypes. *European Journal of Personality, 21*, 953-976. doi:10.1002/per.647

Minkov, M. (2007). *What Makes Us Different and Similar: A New Interpretation of the World Values Survey and Other Cross-Cultural Data*. Sofia, Bulgaria: Klasika I Stil.

Minkov, M., Blagoev, V. & Bond, M. H. (2015). Improving research in the emerging field of cross-cultural sociogenetics: The case of serotonin. *Journal of Cross-Cultural Psychology, 46*, 336-354.

Montirosso R, Cozzi P., Putnam S. P., Gartstein M. A., & Borgatti R. (2011). Studying cross-cultural differences in temperament in the first year of life: United States and Italy. *International Journal of Behavioral Development, 35*, 27-37. doi:10.1177/0165025410368944

Parade, S. H. & Leerkes, E. M. (2008). The reliability and validity of the Infant Behavior Questionnaire-Revised. *Infant Behavior and Development, 31*, 637-646. Doi: 10.1016/j.infbeh.2008.07.009

Peterson, E. R., Waldie, K. E., Mohal, J., Reese, E., Atatoa Carr, P. E., Grant, C. C., Morton, S. M. B. (2017). Infant Behavior Questionnaire Revised–Very Short Form: A new factor structure's associations with parenting perceptions and child language outcomes. *Journal of Personality Assessment*, *99*, 561-573. doi:10.1080/00223891.2017.1287709

Polo, A. J. & Lopez, S. R. (2009). Culture, context and the internalizing distress of Mexican American youth. *Journal of Clinical Child and Adolescent Psychology, 38*, 273-285. doi:10.1080/15374410802698370

Putnam, S. P., Ellis, L. K., & Rothbart, M.K. (2001). The structure of temperament from infancy through adolescence. In A. Eliasz & A. Angleitner (Eds.) *Advances/proceedings in research on temperament* (pp. 165-182). Germany: Pabst Scientist Publisher.

Putnam, S. P. & Gartstein, M. A. (2017). Aggregate temperament scores from 18 countries: Associations with aggregate personality traits, cultural dimensions, and allelic frequency. *Journal of Research in Personality*, *67*, 157-170. doi:10.1016/j.jrp.2016.07.009

Putnam, S. P., Gartstein, M. A., & Rothbart, M. K. (2006). Measurement of fine-grained aspects of toddler temperament: The Early Childhood Behavior Questionnaire. *Infant Behavior and Development, 29*, 386-401. doi:10.1016/j.infbeh.2006.01.004

Putnam, S.P., Helbig, A., Gartstein, M. A., Rothbart, M. K. & Leerkes, E. M. (2014). Development and assessment of short and very short forms of the Infant Behavior Questionnaire-Revised. *Journal of Personality Assessment, 96, 445-458*. doi:10.1080/00223891.2013.841171

Putnam, S. P., Jacobs, J., Gartstein, M. A., & Rothbart, M. K. (2010, March). Development and assessment of short and very short forms of the Early Childhood Behavior Questionnaire. Poster presented at International Conference on Infant Studies, Baltimore, MD.

Putnam, S. P., & Rothbart, M. K. (2006). Development of short and very short forms of the Children's Behavior Questionnaire. *Journal of Personality Assessment, 87*, 102-112. doi:10.1207/s15327752jpa8701_09

Putnam, S. P., Rothbart, M. K., & Gartstein, M. A. (2008). Homotypic and heterotypic continuity of fine-grained temperament during infancy, toddlerhood, and early childhood. *Infant and Child Development, 17*, 387-405. doi:10.1002/icd.582

Raudenbush and Bryk (2002). *Hierarchical Linear Models: Applications and Data Analysis Methods.* University of Chicago Press.

Raval, V. V., Daga, S. S., Raval, P. H. & Panchal, I. N. (2016). Asian Indian mothers' emotion socialization and child emotion expression as a function of situational context. *Journal of Child and Family Studies, 25*, 2853-2861.

Raval, V. V. & Martini, T. S. (2011). Making the child understand: Socialization of emotion in urban India. *Journal of Family Psychology, 25*, 847-856,

Raval, V. V. & Walker, B. L. (2019). Unpacking 'culture': Caregiver socialization of emotion and child functioning in diverse families. *Developmental Review, 51*, 146-174. Doi: 10.1016/j.dr.2018.11.001

Rentfrow, P. J., Gosling, S. D. & Potter, J. (2008). A theory of the emergence, persistence, and expression of geographic variation in psychological characteristics. *Perspectives on Psychological Science, 3(5),* 339-369. doi: 10.1111/j.1745-6924.2008.00084.x

Ross, M. H. & Homer, E. (1976). Galton's problem in cross-national research. *World Politics, 29*, 1-28.

Rothbart, M. K. (1981). Measurement of temperament in infancy. *Child Development, 52*, 569-578. doi:10.2307/1129176

Rothbart, M. K. (2011). *Becoming who we are: Temperament and personality in development*. New York: Guilford Press.

Rothbart, M. K., Ahadi, S. A., Hershey, K. L., & Fisher, P. (2001). Investigations of temperament at three to seven years: The Children's Behavior Questionnaire. *Child Development, 72*, 1394–1408. doi:10.1111/1467-8624.00355

Rothbart, M. K., & Bates, J. E. (2006). Temperament. In W. Damon, R. Lerner, & N. Eisenberg (Eds.), *Handbook of child psychology: Vol. 3. Social, emotional, and personality development* (6th ed., pp. 99–166). New York: Wiley.

Rothbart, M. K. & Derryberry, D. (1981). Development of individual differences in temperament. In M. E. Lamb, & A. L. Brown (Eds.), *Advances in Developmental Psychology* (Vol. 1, pp. 37-86). Hillsdale, NJ: Erlbaum.

Rothbaum, F. & Trommsdorff, G. (2007). Do Roots and Wings Complement or Oppose One Another?: The Socialization of Relatedness and Autonomy in Cultural Context. In J. E. Grusec & P. D. Hastings (Eds.), *Handbook of socialization: Theory and research* (pp. 461–489). The Guilford Press.

Rubin, K. H., Hemphill, S.A., Chen, X., Hastings, P., Sanson, A., Coco, A. L., Zappulla, C., Chung, O.-B., Park, S.-Y., Doh, H. S., Chen, H., Sun, L., Yoon, C. H., & Cui, L. (2006). A cross-cultural study of behavioral inhibition in toddlers: East-West-North-South. *International Journal of Behavioral Development, 30,* 219-226. doi:10.1177/0165025406066723

Schaller, M. & Murray, D.R. (2008). Pathogens, personality and culture: Disease prevalence predicts worldwide variability in sociosexuality, extraversion and openness to experience. *Journal of Personality and Social Psychology, 95*, 212-221. doi:10.1037/0022-3514.95.1.212

Schmitt, D. P., Allik, J., Mccrae, R. R., & Benet-Martínez, V. (2007). The geographic distribution of Big Five personality traits: Patterns and profiles of human self-description across 56 nations. *Journal of Cross-Cultural Psychology, 38*, 173–212. doi:10.1177/0022022106297299

Schmitt, D. P., Realo, A., Voracek, M. & Allik, J. (2008). Why can't a man be more like a woman? Sex differences in Big Five personality traits across 55 cultures. *Journal of Personality and Social Psychology, 94*, 168-182. DOI:10.1037/a0014651

Senese, V. P., Bornstein, M. H., Haynes, O. M., Rossi, G., Venuti, P. A. (2012). Cross-cultural comparison of mothers' beliefs about their parenting very young children. *Infant Behavior and Development, 35*, 479-88. doi: 10.1016/j.infbeh.2012.02.006.

Shiner, R. L., Buss, K. A., McClowry, S. G., Putnam, S. P., Saudino, Z. J., & Zentner, M. (2012). What is temperament now? Assessing progress in temperament research on the twenty-fifth anniversary of Goldsmith et al. (1987). *Child Development Perspectives, 6*, 436–444. doi:10.1111/j.1750-8606.2012.00254.x

Shiner, R. & Caspi, A. (2003). Personality differences in childhood and adolescence: Measurement, development, and consequences. *Journal of Child Psychology and Psychiatry, 44*, 2-32. doi:10.1111/1469-7610.00101

Simonds, J. (2006). *The role of reward sensitivity and response: Execution in childhood extraversion.* Unpublished doctoral dissertation, University of Oregon.

Sleddens, E. F. C., Kremers, S. P. J, Candel, M. J. J. M., De Vries, N. N. K. & Thijs, C. (2011). Validating the Children's Behavior Questionnaire in Dutch children: Psychometric properties and a cross-cultural comparison of factor structures. *Psychological Assessment, 23*, 417-426. doi:10.1037/a0022111

Slobodskaya, H. R., Gartstein, M.A., Nakagawa, A., & Putnam, S. P. (2013). Early temperament in Japan, US and Russia: Do cross-cultural differences decrease with age? *Journal of Cross-Cultural Psychology*, *44*, 438 460. doi:10.1177/0022022112453316

Stępień-Nycz, M., Rostek, I., Białecka-Pikul, M. & Białek, A. (2018). The Polish adaptation of the Early Childhood Behavior Questionnaire (ECBQ): Psychometric properties, age and gender differences and convergence between the questionnaire and the observational data. *European Journal of Developmental Psychology*, *15*, 192-213. doi:10.1080/17405629.2017.1292906

Strickhouser, J. E. & Sutin, A. R. (2020). Family and neighborhood socioeconomic status and temperament development from childhood to adolescence. *Journal of Personality, 88,* 515-529. doi:10.1111/jopy.12507

Super, C. M., Axia, G., Harkness, S., Welles-Nystrom, B., Zylicz, P. O., Parmar, P., & McGurk, H. (2008). Culture, temperament, and the “difficult child”: A study in seven western cultures. *International Journal of Developmental Science*, *2*(1), 136-157. doi:10.3233/DEV-2008-21209

Super, C. M., & Harkness, S. (1986). The developmental niche: A conceptualization at the interface of child and culture. *International Journal of Behavioral Development, 9*, 545-569. doi:10.1177/016502548600900409

Super, C.M., Harkness, S., Bonichini, S., Welles, B., Zylicz, P.O., Bermúdez, M., Rios, P. J. (2020). Developmental continuity and change in the cultural construction of the 'difficult child': A study in six Western cultures. *New Directions for Child and Adolescent Development, 170*, 43-68. doi:10.1002/cad.20338

Terracciano, A., Abdel-Khalak, A. M., A ´ da´m, N., Adamovova´ L., Ahn, C.-k., Ahn, H.-n., et al. (2005). National character does not reflect mean personality trait levels in 49 cultures. *Science,310*, 96–100.

Thomas, A., & Chess, S. (1977). *Temperament and development*. New York: Brunner/Mazel.

Trommsdorff, G., Cole, P. & Heikamp, T. (2012). Cultural variations in mothers' intuitive theories: A preliminary report on interviewing mothers from five nations about their socialization of children's emotions. *Global Studies of Childhood, 2*, 158-169. Doi: 10.2304/gsch.2012.2.2.158

van Dijk, R., van der Valk, I., Dekovic, M. & Branje, S. (2020). A meta-analysis on interparental conflict, parenting, and child adjustment in divorced families: Examining mediation using meta-analytic structural equation models. *Clinical Psychology Review, 79,*101861. doi:10.1016/j.cpr.2020.101861

Vygotsky, L.S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MAS: Harvard University Press.

Way, B. & Lieberman, M. D. (2010). Is there a genetic contribution to cultural differences? Collectivism, individualism, and genetic markers of social sensitivity. *Social Cognitive and Affective Neuroscience, 5*, 203-211.

Zhang, L., Lee, A. J., DeBruine, L. M., & Jones, B. C. (2019). Are sex differences in preferences for physical attractiveness and good earning capacity in potential mates smaller in countries with greater gender equality? *Evolutionary Psychology*, *17*, Article 1474704919852921. doi.org/10.1177/1474704919852921

Footnote

[1] Consistent with Else-Quest et al. (2006), we use the term gender, rather than sex, as reported by parents and presumably assigned at birth.

Table 1. Between- and Within-Nation Standard Deviation Analyses and Nation X Gender ANOVA for IBQ-R, ECBQ and CBQ scales.

	IBQR SUR	IBQR NEG	IBQR REG	ECBQ SUR	ECBQ NEG	ECBQ REG	CBQ SUR	CBQ NEG	CBQ REG
Standard Deviation Analyses									
Mean Within-Nation *SD*	.881	.994	.794	.752	.725	.776	.880	.845	.757
Between-Nation *SD*	.376	.281	.341	.215	.389	.262	.200	.345	.256
Between-Nation Variance Proportion	.182	.080	.184	.082	.290	.114	.049	.166	.114
ANOVA									
Age F	7602.71**	1838.67**	.88	17.49**	155.7**	557.01**	14.22**	224.66**	73.07**
Nation F	38.10**	49.78**	81.79**	17.14**	97.24**	39.92**	37.10**	119.63**	48.52**
Gender F	4.80*	.07	1.88	8.45**	3.00	38.42**	159.91**	15.05**	325.06**
Nation*Gender F	2.06**	1.45*	1.61*	1.02	1.07	1.19	2.21**	1.73**	1.62**
Age Partial Eta^2	.219	.064	.000	.001	.008	.030	.000	.006	.002
Nation Partial Eta^2	.044	.057	.091	.034	.168	.077	.044	.130	.057
Gender Partial Eta^2	.000	.000	.000	.000	.000	.002	.004	.000	.009
Nation*Gender Partial Eta^2	.003	.002	.002	.002	.002	.002	.003	.002	.002

Note. SUR = Surgency. NEG = Negative Affectivity. REG = Regulatory Capacity. Nation and Gender X Nation *df*s = 33 for IBQR tests, 38 for ECBQ tests and 45 for CBQ tests. Error *df*s = 27,051 for IBQ-R, 18,267 for ECBQ, and 36,009 for CBQ. * $p < .05$, ** $p < .01$

Table 2. Estimated Nation-Level Means of Refined Temperament Scales.

Nation	IBQR SUR	IBQR NEG	IBQR REG	ECBQ SUR	ECBQ NEG	ECBQ REG	CBQ SUR	CBQ NEG	CBQ REG	Omnibus SUR	Omnibus NEG	Omnibus REG
Argentina							4.73	4.54	5.72	0.86	0.88	1.47
Australia	4.66	3.62	4.59	5.68	2.81	4.48				0.89	-0.58	-0.54
Belgium	4.59	3.25	5.32	5.11	2.57	4.40	4.27	4.50	5.38	-0.9	-0.74	0.27
Bos & Herz							4.88	4.30	5.71	1.63	0.18	1.43
Brazil				5.25	3.24	4.48	4.45	4.58	4.77	-0.10	1.00	-1.36
Canada	4.85	3.87	5.08	5.24	2.71	4.77	4.72	3.88	5.42	0.37	-0.56	0.65
Chile	4.93	4.03	5.06	5.44	3.32	4.59	4.31	3.93	5.31	0.09	0.19	0.22
China	4.70	3.86	4.34	4.90	3.60	4.34	4.22	4.06	5.17	-1.19	0.38	-1.05
Colombia				4.68	3.90	4.22	4.52	4.91	4.55	-1.28	2.35	-2.37
Curacao							4.35	3.88	5.55	-1.12	-1.03	0.81
Czech Rep.	4.54	3.67	4.76	5.22	3.16	4.59	4.57	3.82	5.35	-0.27	-0.44	-0.02
Denmark				5.27	2.26	4.88				0.51	-1.54	1.37
Estonia				5.12	2.70	4.74				-0.21	-0.41	0.71
Finland	4.82	3.53	4.76	5.21	2.51	4.74	4.61	3.63	5.38	0.10	-1.33	0.23
France				5.11	2.80	4.40				-0.26	-0.13	-0.79
Germany	4.69	3.84	4.47	4.96	2.49	4.32	4.83	4.08	5.25	-0.06	-0.59	-0.84
Greece							4.76	4.34	5.36	1.02	0.28	0.04
Hong Kong							4.35	4.12	5.16	-1.11	-0.35	-0.73
Hungary	4.97	3.61	5.21	5.10	2.83	5.07	4.31	4.16	5.45	-0.40	-0.45	1.25
India							4.60	4.40	5.52	0.18	0.48	0.7
Indonesia	5.35	3.97	4.92	5.42	3.15	5.05	4.71	4.05	5.39	1.23	0.09	0.87
Iran	4.54	4.84	4.12				4.69	4.50	5.66	-0.21	1.78	-0.41
Ireland				5.33	2.34	4.58				0.80	-1.35	0.02
Israel	4.90	3.96	4.98	5.11	2.98	4.72	4.46	4.37	5.55	-0.22	0.24	0.64
Italy	4.74	4.04	4.84	5.12	2.77	4.44	4.70	4.45	5.61	0.02	0.23	0.17
Japan	4.46	4.02	4.22	5.12	2.82	4.17	4.41	4.12	5.12	-0.79	-0.08	-1.49
Kosovo	5.04	4.46	5.02	5.12	3.70	4.85	4.63	4.88	5.42	0.25	1.90	0.71
Latvia	4.79	3.84	4.72							-0.23	-0.35	-0.27
Lithuania				5.15	2.93	4.53	4.27	4.27	5.12	-0.8	0.16	-0.55
Malaysia	5.42	4.19	5.21				4.32	4.92	5.24	0.36	1.37	0.36
Malta	5.34	4.06	5.48							1.71	0.34	1.96

Mexico	5.15	4.09	5.21	5.11	2.58	4.78	4.35	4.16	5.03	-0.10	-0.16	0.27
Myanmar							4.82	4.42	5.56	1.31	0.52	0.84
Netherlands	4.32	3.29	4.76	5.04	2.45	4.70	4.57	3.37	5.01	-0.80	-1.89	-0.32
New Zealand	4.93	3.80	4.48				4.59	4.04	5.53	0.20	-0.54	-0.12
Nigeria	4.99	4.31	5.09	5.30	3.38	4.43				0.56	1.26	0.08
Pakistan							4.95	4.82	5.52	1.96	1.69	0.69
Peru							4.54	4.32	5.61	-0.14	0.23	1.06
Philippines							4.73	4.75	5.58	0.84	1.50	0.94
Poland	4.53	4.09	4.48	4.91	2.79	4.39	4.56	4.26	5.40	-0.78	0.11	-0.53
Portugal	4.97	4.19	5.13	5.34	3.30	4.68	4.57	4.35	5.31	0.42	0.76	0.41
Romania	5.26	4.16	5.18	5.14	2.62	4.58	4.43	3.92	4.92	0.21	-0.29	-0.19
Russia	4.57	4.10	4.43	5.05	2.60	4.48				-0.76	-0.1	-0.78
Serbia							4.58	4.03	5.61	0.08	-0.60	1.03
Singapore							4.53	4.26	5.33	-0.17	0.06	-0.06
Slovakia							4.93	4.19	5.20	1.89	-0.14	-0.56
Slovenia				5.31	2.57	4.64				0.66	-0.74	0.28
S. Korea	5.02	4.14	4.88	5.12	2.75	4.88	4.36	4.02	5.30	-0.23	-0.10	0.45
Spain	4.91	3.92	4.85	5.26	2.68	4.50	4.31	3.98	5.11	-0.23	-0.43	-0.40
Suriname							4.86	4.17	5.70	1.51	-0.19	1.40
Sweden	4.45	3.51	4.36	4.96	2.29	4.69				-1.17	-1.44	-0.41
Switzerland				4.87	2.39	4.46	4.72	3.71	5.61	-0.29	-1.37	0.26
Taiwan	4.79	3.99	4.42	4.94	3.03	4.51	4.39	4.17	5.30	-0.72	0.12	-0.55
Thailand				5.67	2.93	4.96	4.46	4.19	4.98	0.93	0.02	0.13
Türkiye	5.37	4.28	5.11	4.83	2.98	4.17	4.69	4.92	5.54	0.28	1.12	-0.07
UK	4.55	3.88	4.74	5.45	2.88	4.47	4.63	4.26	5.34	0.20	-0.03	-0.25
Ukraine				5.39	2.63	4.41				1.06	-0.57	-0.76
Uruguay	4.85	4.10	4.59							-0.02	0.47	-0.66
US	4.95	3.90	4.86	5.11	2.76	4.43	4.76	3.96	5.29	0.36	-0.40	-0.25

Note. SUR = Surgency. NEG = Negative Affectivity. REG = Regulatory Capacity. Marginal means represent scores from revised scales resulting from measurement invariance analyses, corrected for gender and age. IBQ-R means standardized for 7.97 months. ECBQ means standardized for 25.23 months. CBQ means standardized for 61.44 months. OMNI = Omnibus scores are the average of standardized (z-scores) marginal means. Bos & Herz = Bosnia and Herzegovina. UK = United Kingdom. US = United States.

Table 3. Estimated Nation-Level Means of Original Temperament Scales.

Nation	IBQR SUR	IBQR NEG	IBQR REG	ECBQ SUR	ECBQ NEG	ECBQ REG	CBQ SUR	CBQ NEG	CBQ REG
Argentina							4.73	4.54	4.73
Australia	4.64	3.45	4.79	5.57	2.92	4.60			
Belgium	4.50	3.35	5.45	5.07	2.60	4.57	4.27	4.50	4.16
Bos & Herz							4.88	4.30	4.72
Brazil				5.09	3.37	4.68	4.45	4.58	4.51
Canada	4.75	3.82	5.25	5.10	2.76	4.88	4.72	3.88	4.53
Chile	4.88	3.89	5.24	5.26	3.34	4.78	4.31	3.93	4.38
China	4.55	3.72	4.52	4.87	3.65	4.57	4.22	4.06	4.14
Colombia				4.72	3.93	4.44	4.52	4.91	4.41
Curacao							4.35	3.88	4.32
Czech Rep.	4.41	3.56	4.97	5.05	3.18	4.69	4.57	3.82	4.40
Denmark				5.19	2.33	4.97			
Estonia				5.17	2.81	4.81			
Finland	4.64	3.35	4.99	5.10	2.53	4.87	4.61	3.63	4.54
France				4.84	2.99	4.46			
Germany	4.54	3.58	4.72	4.82	2.52	4.52	4.83	4.08	4.77
Greece							4.76	4.34	4.56
Hong Kong							4.35	4.12	4.28
Hungary	4.88	3.46	5.42	5.12	2.92	5.16	4.31	4.16	4.28
India							4.60	4.40	4.58
Indonesia	5.32	3.82	5.09	5.24	3.10	5.18	4.71	4.05	4.44
Iran	4.36	4.60	4.35				4.69	4.50	4.63
Ireland				5.27	2.41	4.67			
Israel	4.81	3.82	5.14	5.07	3.01	4.79	4.46	4.37	4.37
Italy	4.67	3.85	5.04	5.05	2.81	4.61	4.70	4.45	4.61
Japan	4.40	3.89	4.51	4.89	2.86	4.48	4.41	4.12	4.25
Kosovo	4.98	4.39	5.17	5.07	3.85	4.87	4.63	4.88	4.58
Latvia	4.58	3.69	4.95						
Lithuania				5.02	2.93	4.70	4.27	4.27	4.37
Malaysia	5.25	4.09	5.35				4.32	4.92	4.19
Malta	5.25	3.93	5.61						
Mexico	5.16	3.86	5.38	4.83	2.62	4.87	4.35	4.16	4.35

Myanmar							4.82	4.42	4.69
Netherlands	4.32	3.13	4.99	5.02	2.43	4.75	4.57	3.37	4.46
New Zealand	4.82	3.73	4.69				4.59	4.04	4.34
Nigeria	4.90	4.29	5.22	5.10	3.42	4.69			
Pakistan							4.95	4.82	4.85
Peru							4.54	4.32	4.47
Philippines							4.73	4.75	4.62
Poland	4.51	3.87	4.65	4.81	2.76	4.55	4.56	4.26	4.55
Portugal	4.96	4.09	5.22	5.25	3.37	4.81	4.57	4.35	4.50
Romania	5.23	4.02	5.31	5.00	2.66	4.76	4.43	3.92	4.43
Russia	4.50	3.89	4.59	4.93	2.71	4.60			
Serbia							4.58	4.03	4.59
Singapore							4.53	4.26	4.45
Slovakia							4.93	4.19	4.77
Slovenia				5.23	2.57	4.71			
S. Korea	4.97	3.96	5.06	4.93	2.94	5.01	4.36	4.02	4.26
Spain	4.87	3.72	5.04	4.94	2.79	4.60	4.31	3.98	4.37
Suriname							4.86	4.17	4.77
Sweden	4.41	3.46	4.61	4.78	2.23	4.73			
Switzerland				4.67	2.49	4.65	4.72	3.71	4.53
Taiwan	4.65	3.77	4.70	4.72	3.04	4.65	4.39	4.17	4.30
Thailand				5.48	2.98	5.12	4.46	4.19	4.43
Türkiye	5.24	4.16	5.24	4.63	2.93	4.40	4.69	4.92	4.58
UK	4.50	3.79	4.93	5.33	2.94	4.54	4.63	4.26	4.42
Ukraine				5.22	2.70	4.56			
Uruguay	4.86	3.89	4.83						
US	4.90	3.75	5.05	5.04	2.82	4.54	4.76	3.96	4.65

Note. SUR = Surgency. NEG = Negative Affectivity. REG = Regulatory Capacity. Marginal means represent scores from original IBQ-R, ECBQ and CBQ scales, corrected for gender and age. IBQ-R means standardized for 7.97 months. ECBQ means standardized for 25.23 months. CBQ means standardized for 61.44 months. OMNI = Omnibus scores are the average of standardized (z-scores) marginal means. Bos & Herz = Bosnia and Herzegovina. UK = United Kingdom. US = United States.

Table 4. Cross-age Correlations between IBQ-R, ECBQ and CBQ Scales.

		ECBQ			CBQ		
		SUR	NEG	REG	SUR	NEG	REG
	SUR	.10	.32	.32#	-.08	.39*	-.02
IBQ-R	NEG	-.04	.53**	-.09	.17	.57*	.24
	ORC	.25	.13	.41*	-.21	.21	-.11
	SUR				-.07	-.22	.02
ECBQ	NEG				-.17	.58**	-.28
	EFF				-.06	-.28	.16

Note. $n = 28$ for IBQ-R to ECBQ, $n = 27$ for ECBQ to CBQ, $n = 29$ for IBQ-R to CBQ. SUR = Surgency. SUR = Surgency. NEG = Negative Affectivity. REG = Regulatory Capacity. * $p < .05$, ** $p < .01$.

Table 5. Correlations between Aggregate Personality and Aggregate Temperament scores.

	Surgency	Negative Affectivity	Regulatory Capacity
Extraversion			
Self	.39*	-.26	.15
Other	.21	-.21	.16
Neuroticism			
Self	-.38*	.33*	-.11
Other	.11	.18	.04
Conscientiousness			
Self	.54**	.24	.24
Other	.08	-.03	.18
Openness to Experience			
Self	.01	-.28#	-.05
Other	-.17	-.46**	-.12
Agreeableness			
Self	.15	-.09	.08
Other	.11	-.25	.06

Note: Correlation n = 37 for Self-Reported Personality, n = 27 for Other-Reported Personality

** $p < .10$. * $p < .05$. # $p < .10$.

Table 6. Correlations and Regression Coefficients between Cultural Orientation Dimensions and Aggregate Temperament scores.

	Surgency		Negative Affectivity		Regulatory Capacity	
	r	β	*r*	β	*r*	β
Individualism	-.05	.01	-.56**	-.29*	-.12	.01
Power Distance	.14	.03	.45**	-.06	-.07	-.17
Masculinity	.13	.17	.20	.26*	-.13	-.07
Uncertainty Avoidance	.05	-.03	.21	-.01	-.01	.01
Indulgence	-.02	-.28	-.10	-.03	.11	-.10
Long Term Orientation	-.24#	-.46*	-.33*	-.23	-.19	-.21
GNI per capita	-.25#	.01	-.65**	-.50**	-.14	-.19
F		1.32		9.10**		.49
R^2		.18		.60		.07

Note: Correlation *n* = 53 for Individualism, Power Distance, Masculinity, and Uncertainty Avoidance; *n* = 55 for Indulgence; *n* = 54 for Long-Term Orientation; *n* = 59 for GNI-PC.

Regression *df* = 7, 43.

** $p < .10$. * $p < .05$. # $p < .10$.

Table 7. UN Region X Gender ANOVA and Regional Marginal Means for Refined IBQ-R, ECBQ and CBQ scales.

Nation	IBQR SUR	IBQR NEG	IBQR REG	ECBQ SUR	ECBQ NEG	ECBQ REG	CBQ SUR	CBQ NEG	CBQ REG	Omnibus SUR	Omnibus NEG	Omnibus REG
ANOVA												
Age F	8351.85**	1718.68**	35.49**	65.58**	80.79**	1099.27**	1.68	112.44**	32.63**			
Region F	48.89**	102.23**	119.91**	28.58**	193.04**	32.10**	94.86**	299.64**	64.97**			
Gender F	4.33*	1.09	0.51	8.16**	0.96	25.40**	85.55**	8.85**	132.56**			
Region*Gender F	1.79*	2.20**	1.59	1.31	0.90	1.50	1.51	2.10*	2.19**			
Age Partial Eta^2	.236	.060	.001	.004	.004	.057	.000	.003	.001			
Region Partial Eta^2	.021	.043	.050	.017	.104	.019	.031	.091	.021			
Gender Partial Eta^2	.000	.000	.000	.000	.000	.001	.002	.000	.004			
Region*Gender Partial Eta^2	.001	.001	.001	.001	.001	.001	.001	.001	.001			
Marginal Means												
Australia/NZ	4.79	3.71	4.53	5.66	2.84	4.44	4.58	4.02	5.52	0.67	-0.39	-0.36
Caribbean (Curacao)							4.35	3.90	5.56	-1.26	-0.66	1.29
Central America (Mexico)	5.16	4.09	5.22	5.10	2.59	4.77	4.35	4.16	5.03	-0.22	-0.17	0.11
Eastern Asia	4.72	3.95	4.42	5.01	3.13	4.48	4.30	4.07	5.19	-1.10	0.19	-1.05
Eastern Europe	4.77	3.74	4.94	5.09	2.91	4.75	4.47	4.00	5.22	-0.49	-0.32	0.14
Northern America	4.90	3.88	4.98	5.20	2.72	4.67	4.73	3.90	5.36	0.45	-0.48	0.29
Northern Europe	4.81	3.55	4.76	5.09	2.38	4.72	4.59	3.70	5.36	-0.17	-1.35	0.15
South America	4.89	4.06	4.89	5.32	3.34	4.51	4.54	4.32	5.37	0.21	0.77	-0.11
Southeastern Asia	5.40	4.09	5.10	5.59	2.99	4.98	4.55	4.42	5.29	1.34	0.52	0.92
Southern Asia	4.55	4.82	4.15	5.22	2.84	4.54	4.79	4.63	5.55	0.16	1.27	-0.53
Southern Europe	4.86	4.06	4.94	5.30	3.39	4.42	4.50	4.20	5.28	0.00	0.70	-0.42
Sub-Saharan Africa (Nigeria)	4.99	4.32	5.07							0.47	0.95	0.79
Western Asia	5.04	4.06	5.02	5.09	2.98	4.67	4.57	4.64	5.55	0.11	0.70	0.69
Western Europe	4.50	3.43	4.85	5.02	2.48	4.56	4.60	3.53	5.13	-0.69	-1.54	-0.53

Note: SUR = Surgency. NEG = Negative Affectivity. REG = Regulatory Capacity. Region and Gender X Region dfs = 12 for IBQR tests, 11 for ECBQ tests and 12 for CBQ tests. Error dfs = 27,093 for IBQ-R, 18,321 for ECBQ, and 36,075 for CBQ. * $p < .05$, ** $p < .01$. Marginal means represent scores from revised scales resulting from measurement invariance analyses, corrected for gender and age. IBQ-R means standardized for 7.97 months. ECBQ means standardized for 25.23 months. CBQ means standardized for 61.44 months.

Figure 1. Geographic Distribution of Aggregate Surgency Scores. Darker shading represents higher scores. Data not available for countries in white.

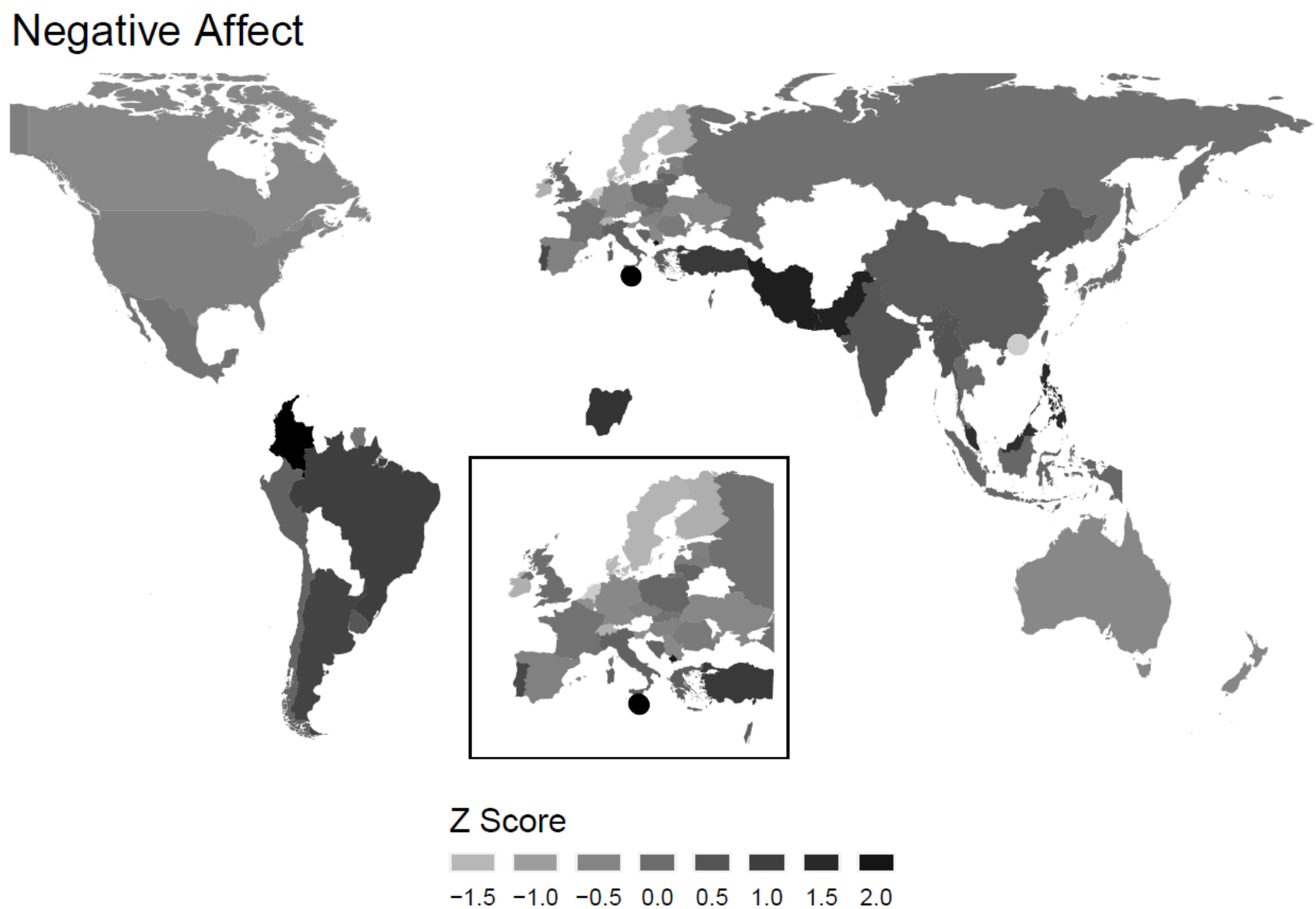

Figure 2. Geographic Distribution of Aggregate Negative Affectivity Scores. Darker shading represents higher scores. Data not available for countries in white.

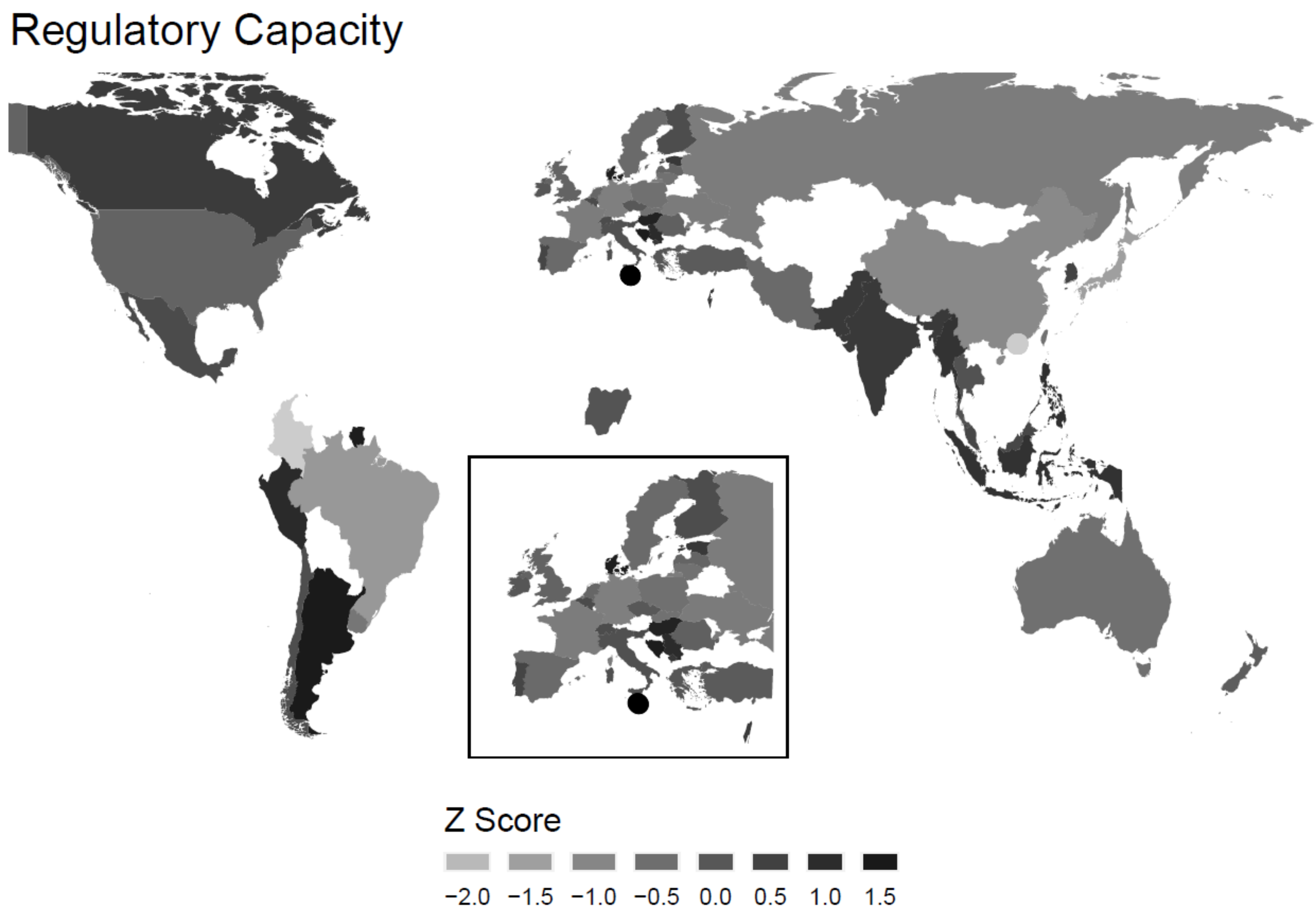

Figure 3. Geographic Distribution of Aggregate Regulatory Capacity Scores. Darker shading represents higher scores. Data not available for countries in white.

Appendix 1

Contributing members to the Global Temperament Project.

Last Name	First Name	Middle Name	Institution
Abdul Razak	Nurliyana		Universiti Putra Malaysia
Abela	Angela		University of Malta
Abramson	Lior		The Hebrew University of Jerusalem
Acar	Ibrahim	H.	Department of Psychology, Ozyegin University
Ahmetoglu	Emine		Trakya University
Akhter	Syeda	Rezwana	BRAC University
Aldrete-Cortez	Vania		School of Psychology, Universidad Panamericana, and Organization Escucha con Respeto a tu Bebé
Alexander	Ava		Bowdoin College
Alonso-Alberca	Natalia		University of the Basque Country
Alves	Diana		University of Porto
Amaro-Hinojosa	Marily	Daniela Amaro	Universidad Autónoma de Chihuahua
Amersfoort	Linde-Marie		University of Canterbury
Anson	Elizabeth		University of Rochester
Arck	Petra		Department of Obstetrics and Fetal Medicine, University Medical Center Hamburg-Eppendorf
Armony-Sivan	Rinat		Ashkelon Academic College
Aschersleben	Gisa		Saarland University
Assor	Avi		Ben Gurion University
Atzaba-Poria	Naama		Ben-Gurion University of the Negev
Aureli	Tiziana		University of Chieti-Pescara
Aydın Uysal	Ayse		Anadolu University
Ayneto	Alba		Universitat Pompeu Fabra
Bachner-Melman	Rachel		Ruppin Academic Center and the Hebrew University of Jerusalem
Bae	Yun-Jin		Korea Institute of Child Care and Education
Bai	Liu		Beijing Normal University
Baillet	Thalia	Harmony	Universidad Nacional Autónoma de México
Bajgarová	Zdeňka		University of South Bohemia
Bangpan	Laisen		Department of Pediatrics, Siriraj Hospital, Mahidol University
Baranauskienė	Ingrida		Lithuanian University of Educational Sciences
Barbosa	Miguel		Centro de Investigação em Ciência Psicológica, Faculdade de Psicologia, Universidade de Lisboa
Barcaccia	Barbara		Sapienza University of Rome
Barisnikov	Koviljka		Université de Genève
Bárrig Jó	Patricia		Universidad de San Martin de Porres
Barros	Luisa		Centro de Investigação em Ciência Psicológica, Faculdade de Psicologia, Universidade de Lisboa
Bassan	Haim		Shamir Medical Center, Tel Aviv University
Bates	John	E.	Indiana University
Baumgartner	Emma		Sapienza University of Rome
Beijers	Roseriet		Radboud University, Radboud University Medical Center
Bell	Martha Ann		Virginia Tech
Benga	Oana		Babes Bolyai University
Berger	Andrea		Ben-Gurion University of the Negev

Last Name	First Name	Middle Name	Institution
Bernier	Annie		University of Montreal
Białecka-Pikul	Marta		Jagiellonian University
Birken	Catherine	S.	Hospital for Sick Children, University of Toronto
Birle	Delia		University of Oradea
Blazkova	Iveta		University of South Bohemia
Blissett	Jacqueline		Aston University
Bögels	Susan		University of Amsterdam
Bonichini	Sabrina		University of Padua
Boothroyd	Lynda	G.	Durham University
Bornstein	Marc	H.	Eunice Kennedy Shriver National Institute of Child Health and Human Development, IFS, UNICEF
Borovnica	Olja		Clinical Hospital Centre "Zvezdara", ENT Clinic, Center for early diagnostic and therapy for children with hearing loss
Boterberg	Sofie		Ghent University - Department of Experimental-Clinical and Health Psychology
Bozicevic	Laura		Institute of Population Health, Department of Primary Care & Mental Health, Faculty of Health and Life Sciences, University of Liverpool
Braeken	Marijke		Rehabilitation Research Center, Biomedical Research Institute, Faculty of Rehabilitation Sciences, Hasselt
Brand	Rebecca		Villanova University
Braungart-Rieker	Julie		Colorado State University, Department of Human Development and Family Studies
Bridgett	David		Northern Illinois University
Buist	Kirsten	L.	Utrecht University
Bureau	Jean-Francois		University of Ottawa
Buss	Kristin	A.	The Pennsylvania State University
Butti	Niccolò		0-3 Centre for the at-Risk Infant, Scientific Institute, IRCCS Eugenio Medea
Cadima	Joana		University of Porto
Cadime	Irene		Universidade do Minho
Čagran	Branka		University of Maribor
Calkins	Susan		University of North Carolina-Greensboro
Canário	Catarina		University of Porto
Carlson	Stephanie		University of Minnesota--Twin Cities
Carrasco	Miguel Á.		UNED, Universidad Nacional de Educación a Distancia
Casalin	Sara		University of Leuven
Cassé	Julie	F. H.	Clinical Child and Family Studies and Amsterdam Public Health research institute, Vrije Universiteit Amsterdam
Castagna	Annalisa		0-3 Centre for the at-Risk Infant, Scientific Institute, IRCCS Eugenio Medea
Castillo	Karen		National Scientific and Technical Research Council (CONICET) and Human, Social and Environmental Science Institute (INCIHUSA)
Chen	Chen		School of Psychology, Nanjing Normal University
Chen	Shing-Jen		Koen Gakuen Women's Junior College
Chen	Yi-Chieh		National Taiwan University
Chen	Yinghe		Beijing Normal University
Chen	Yunn-Ru		Chung Yuan Christian University
Chhangur	Rabia		Tilburg University
Chiang	Wei-Chun		National Taiwan University

Last Name	First Name	Middle Name	Institution
Chien	Hsin-Man		Taipei Tzuchi Hospital
Chukkali	Surekha		Department of Psychology, Christ University Delhi-NCR
Chung	Mira		Sesalmaul SSK Research Center, Gachon University
Chutabhakdikul	Nuanchan		Research Center for Neuroscience, Institute of Molecular Biosciences, Mahidol University, Nakhon Pathom, Thailand
Cieciuch	Jan		Institute of Psychology, Cardinal Stefan Wyszynski University in Warsaw, Poland
Clark	David	Angus	University of Michigan
Cohen	Dina		Oranim Academic College of Education
Cole	Robert		University of Rochester
Colonnesi	Cristina		University of Amsterdam
Conejero	Angela		Universidad de Granada
Conte	Elisabetta		“R. Massa” Department of Human Sciences for Education, University of Milano-Bicocca
Conte	Stefania		State University of New York (SUNY) at Binghamton
Cowan	Phil		University of California, Berkeley
Cowan	Carolyn		University of California, Berkeley
Crnic	Keith	A.	Department of Psychology, Arizona State University
Cruz	Orlanda		University of Porto
Cruz-Alaniz	Yuria		Universidad Nacional Autónoma de México
Csákvári	Judit		Eötvös Loránd University
Danis	Ildiko		Semmelweis University
Daňsová	Petra		Masaryk University
Davidov	Maayan		The Hebrew University of Jerusalem
Davis	Deborah	Winders	University of Louisville School of Medicine and Norton Children Research Institute affiliated with the U of L School of Medicine
de Almeida	Ana	Tomás	Instituto de Educação, University of Minho
de Jong	Marjanneke		Utrecht University
de Klerk	Carina	C.J.M.	Centre for Brain and Cognitive Development, Birkbeck College, University of London; University of Essex
de la Osa	Nuria		Universitat Autonoma de Barcelona
de Vente	Wieke		University of Amsterdam
de Weerth	Carolina		Radboud University Medical Center
Deater-Deckard	Kirby		University of Massachusetts-Amherst
Deković	Maja		Utrecht University
Demirpençe Secinti	Dilara		Sisli Hamidiye Etfal Training and Research Hospital
Demurie	Ellen		Ghent University - Department of Experimental-Clinical and Health Psychology
Dermitzaki	Irini		Department of Early Childhood Education, University of Thessaly, Argonafton & Filellinon
Déry	Michèle		University of Sherbrooke
Dewey	Deborah		University of Calgary
Dias	Paulo		Universidade Católica Portuguesa
Dias	Tatiane	Lebre	Federal University of Mato Grosso
Diemert	Anke		Department of Obstetrics and Fetal Medicine, University Medical Center Hamburg-Eppendorf
DiMaggio	Rosanna		University of Palermo
Ding	Chendong		Nagoya University (Japan) and Northeast Normal University (China)

Last Name	First Name	Middle Name	Institution
Dixon	Wallace	E.	East Tennessee State University
D'Ocon	Ana	Maria	University of Valencia
Donnellan	Brent		Michigan State University
Dotto	Fulvia		Università degli Studi di Trieste
Dragan	Wojciech	L.	Jagiellonian University in Kraków (Poland)
Dubas	Judith	Semon	Utrecht University
Dunsmore	Julie		Virginia Tech
Durbin	Emily		Michigan State University
Dutra	Natália		Durham University
Dyagileva	Yulia		V.I. Vernadsky Crimean Federal University
Ellis	Ashley		Bowdoin College
Enumo	Sônia	Regina Fiorim	Pontifical Catholic University of Campinas
Escalante	Elsa		Universidad del Norte
Esmaeili	Fattaneh		Institute for Cognitive Science Studies
Ezpeleta	Lourdes		Universitat Autonoma de Barcelona
Falck-Ytter	Terje		Uppsala University
Falcón	Alberto		Universidad Autónoma del Estado de Morelos
Farbiash	Tali		Ben-Gurion University of the Negev
Farkas	Chamarrita		Pontificia Universidad Católica de Chile
Farroni	Teresa		Università degli Studi di Padova
Favez	Nicolas		University of Geneva
Fazlıoğlu	Müyesser		Social Science University of Ankara
Ferenczi	Szilvia	Cs.	Apor Vilmos Catholic College
Ferentzi	Hannah		Department of Congenital Heart Disease—Pediatric Cardiology, German Heart Center Berlin
Ferreira	Tiago		Universidade do Porto
Fewtrell	Mary		GOS Institute of Child Health, University College
Fisher	Phillip		University of Oregon
Flores-Peña	Yolanda		Universidad Autónoma de Nuevo León, Facultad de Enfermería
Flynn	Emma		Durham University
Fountoglou	Marina		University of Glasgow
Fraser	William	D.	Université de Sherbrooke
French	Brian	F.	Washington State University
Frenkel	Tahl	I.	Baruch Ivcher School of Psychology, Reichman University, Israel
Friesen	Myron		University of Canterbury
Frigerio	Alessandra		Scientific Institute, IRCCS E. Medea, Child Psychopathology Unit, Bosisio Parini, Lecco, Italy
Galatsopoulou	Katerina		Department of Early Childhood Education, University of Thessaly, Argonafton & Filellinon
Galdiolo	Sarah		University of Mons
Galvis	Esteban		Universidad de los Andes
Ganea	Patricia		University of Toronto
Gardani	Maria		University of Edinburgh
Gartstein	Maria	A.	Washington State University
Gervai	Judit		Institute of Cognitive Neuroscience and Psychology, HUN-REN Research Centre for Natural Sciences, Budapest
Giesbrecht	Gerald		University of Calgary
Gillet	Sophie		University of Liège

Last Name	First Name	Middle Name	Institution
Gliga	Teodora		Birkbeck University
Goes	Ana	Rita	NOVA National School of Public Health, Public Health Research Centre, Comprehensive Health Research Center, CHRC, NOVA University Lisbon, Lisbon, Portugal
Gomes	Ana	Isabel	Centro de Investigação em Ciência Psicológica, Faculdade de Psicologia, Universidade de Lisboa
Gonzales	Nancy	A.	Department of Psychology, Arizona State University
González-Peña	Paloma		UNED, Universidad Nacional de Educación a Distancia
Gonzalez-Salinas	Carmen		University of Murcia
Gosztyla	Dorota		Warsaw School of Social Psychology
Goupil	Louise		PSL Research University
Graham	Susan		University of Calgary
Grech Lanfranco	Ingrid		University of Malta
Greco	Carolina		National Scientific and Technical Research Council
Grossmann	Tobias		University of Virginia and Max Planck Institute
Gunnar	Megan		University of Minnesota
Guo	Fei		Chinese Academy of Sciences
Gur-Yaish	Nurit		Oranim Academic College of Education
Hagenbeek	Sara		University of Amsterdam
Hakim-Larson	Julie		University of Windsor
Han	Sae-Young		Ewha Womans University
Hartman	Catharina	A.	University of Groningen, University Medical Center Groningen
He	Jie		He Jie Zhejiang University
Heilman	Keri	Jeanene	University of North Carolina-Chapel Hill
Heinonen	Kati		Tampere University
Henderson	Annette	M. E.	The University of Auckland
Henning	Anne		SRH Hochschule für Gesundheit
Hildayani	Rini		Universitas Indonesia
Hoicka	Elena		University of Bristol
Holgado-Tello	Francisco-Pablo		UNED, Universidad Nacional de Educación a Distancia
Hong	Ryan	Y.	National University of Singapore
Horodynski	Mildred		Michigan State University
Hoshi	Nobuko		Junior College of Sapporo Otani University
Hosseini	Faezeh	S.	Alzahra University of Tehran
Hoyo	Angela		Universidad de Granada
Hsu	Kung-Yu		National Chung-Cheng University
Huang	Su-Ying		Fu Jen Catholic University
Huitrón	Blanca		National Autonomous University of Mexico
Hutteman	Roos		Utrecht University
Ionio	Chiara		Università Cattolica del Sacro Cuore
Ison	Mirta	Susana	National Scientific and Technical Research Council
Jami	Humaira		Quaid-i-Azam University
Jensen	Lynn		Physiotherapy Department, Perth Children’s Hospital, Perth, Australia
Junttila	Niina		University of Turku
Kabha	Lena		Ben-Gurion University of the Negev
Kallia	Eleni		Department of Early Childhood Education, University of Thessaly, Argonafton & Filellinon
Karagianni	Eleni		University of Aegean

Last Name	First Name	Middle Name	Institution
Karlsson	Hasse		University of Turku
Karlsson	Linnea		University of Turku
Kassir-Erlich	Neta		The Hebrew University of Jerusalem
Kawashima	Akiko		University of Yamanashi
Khaing	Yu Yu		University of Mandalay
Khoury	Manal		University of Haifa
Kılıç	Şükran		Aksaray University, Department of Early Childhood and Education
Kim	Hyoun	K.	Yonsei University
Kim	Jaehee		Ewha Womans University
Kim	Joo Hyun		Graduate School of Korea University
Kim	Min Jeong		Sesalmaul SSK Research Center, Gachon University
Kim	Yoo Jung		Kyung Hee University
Kim	Yunhee		National Youth Policy Institute
Kinsht	Irina		Research Institute of Physiology and Basic Medicine
Knafo-Noam	Ariel		The Hebrew University of Jerusalem
Kochanska	Grazyna		University of Iowa
Koo	Ja Yeun		Yonsei University
Kovač	Jernej		University of Maribor
Kowal	Christine		The Hospital for Sick Children
Kozlova	Elena		Research Institute of Physiology and Basic Medicine
Krauss	Isabelle		Integrated Research Network in Perinatality of Quebec and Eastern Ontario (IRNPQEO)
Kremers	Stef	P.J.	Maastricht University; KOALA Birth Cohort Study, The Netherlands.
Kristen-Antonow	Susanne		Ludwig-Maximilians-University
Krogh	Marianne	T.	Center for Early Intervention and Family Studies, Department of Psychology, University of Copenhagen
Krol	Kathleen		University of Virginia and Max Planck Institute
Kujala	Teija		Centre of Excellence in Music, Mind, Body and Brain Department of Psychology and Logopedics Faculty of Medicine University of Helsinki
Kung	Mei-Juan		Deh Yu College of Nursing and Health
Kunseler	Florentina	C.	Clinical Child and Family Studies and Amsterdam Public Health research institute, Vrije Universiteit Amsterdam
Kusanagi	Emiko		Kokugakuin University Hokkaido Junior College
Kwak	Keumjoo		Seoul National University
Kylliäinen	Anneli		Tampere University
Kyu	Nilar		University of Yangon
Laceulle	Odilia	M.	Utrecht University
Lahti-Pulkkinen	Marius		Department of Psychology and Logopedics, University of Helsinki and Helsinki University Central Hospital
Lakatos	Krisztina		Institute of Cognitive Neuroscience and Psychology, HUN-REN Research Centre for Natural Sciences, Budapest
Lakić	Siniša		University of Banja Luka, Faculty of Philosophy
Lam	Hio	Wa	University of Macau
Lamb	Michael		University of Cambridge
Langerock	Naomi		Université de Genève
Latzman	Robert	D.	Georgia State University
Lay	Keng-Ling		National Taiwan University

Last Name	First Name	Middle Name	Institution
Lecannelier	Felipe		Faculty of Medical Sciences, Universidad de Santiago de Chile
Lee	Eun Gyoung		Ewha Womans University
Lee	Mina		Sesalmaul SSK Research Center, Gachon University
Leerkes	Esther	M.	University of North Carolina-Greensboro
Lejeune	Fleur		Université de Genève
LeMare	Lucy		Simon Fraser University
Lemelin	Jean-Pascal		University of Sherbrooke
Lengua	Liliana		University of Washington
Leppänen	Jukka	M.	University of Turku
Lessing	Nora		University of Hildesheim
Letourneau	Nicole		University of Calgary
Leve	Leslie		University of Oregon
Li	Ishien		National Taichung University of Education
Li	Longfeng		Beijing Normal University
Li	Ming		Peking University First Hospital
Li	Xiaowei		Beijing Normal University
Liang	Zongbao		Southeast University
Lickenbrock	Diane	M.	Western Kentucky University
Lim	Ji-Young		Kyungpook National University
Lin	Betty		Department of Psychology, University at Albany, State University of New York
Lin	Miao-Hui		National University of Tainan
Lin	Qinyi		Beijing Normal University
Linhares	Beatriz		University São Paulo
Lipina	Sebastian		Unidad de Neurobiología Aplicada, CEMIC-CONICET
Lira Luttges	Benjamin		University of Pennsylvania
Lisson	Sveta		Ben-Gurion University of the Negev
Lonigro	Antonia		Sapienza University of Rome
Loop	Laurie		Psychological Sciences Research Institute of the University of Louvain
Lozoff	Betsy		University of Michigan
Luecken	Linda	J.	Department of Psychology, Arizona State University
Luke	Felix		The Hong Kong Polytechnic University
Lv	Pei-Hong		Zhejiang Normal University
Macci Cassia	Viola		University of Milano-Bicocca
Maguire	Jonathon	L.	The Hospital for Sick Children, Dalla Lana School of Public Health, University of Toronto, St. Michael's Hospital
Majdandžić	Mirjana		University of Amsterdam
Majerčáková Albertová	Silvia		Pan-European University
Makhin	Sergei		V.I. Vernadsky Crimean Federal University
Mankuta	David		Hadassah Hospital Obstetrics and Gynecology Department, Hebrew University Medical School
Marquis	Alexandra	R.	'Toronto Metropolitan University
Martinac Dorčić	Tamara		University of Rijeka
Martins	Silvana		Instituto de Educação, University of Minho
Mascheroni	Eleonora		0-3 Centre for the at-Risk Infant, Scientific Institute, IRCCS Eugenio Medea
Masduki	Masadliahani		Universiti Tun Hussein Onn Malaysia
Masopustová	Zuzana		Masaryk University

Last Name	First Name	Middle Name	Institution
Matos	Paula	Mena	Universidade do Porto
Matte-Gagné	Célia		Université Laval
McDonald	Sheila		Alberta Health Services
McFarlane	Natricha	Levy	Hospital for Sick Children, Toronto
McQuillan	Maureen	E.	Indiana University School of Medicine
Menesini	Ersilia		University of Florence
Messinger	Daniel		University of Miami
Mikhailova	Anna		V.I. Vernadsky Crimean Federal University
Mink	Daniela		Saarland University
Miščević	Snježana		University of Banja Luka
Mizuno	Rie		Chukyo University
Mohd Shariff	Zalilah		Universiti Putra Malaysia
Mohd Shukri	Nurul	Husna	Universiti Putra Malaysia
Molfese	Dennis	L.	University of Nebraska-Lincoln
Molfese	Victoria	J.	University of Nebraska-Lincoln
Möller	Corina		Saarland University
Möller	Eline		University of Amsterdam
Monteiro	Lígia		Iscte-Instituto Universitário de Lisboa
Montirosso	Rosario		0-3 Centre for the at-Risk Infant, Scientific Institute, IRCCS Eugenio Medea
Moreira	João		Centro de Investigação em Ciência Psicológica, Faculdade de Psicologia, Universidade de Lisboa
Moreira	Paulo	A. S.	University of Trás-os-Montes and Alto Douro (UTAD), Vila Real, Portugal
Moscardino	Ughetta		University of Padua
Moulson	Margaret	C.	'Toronto Metropolitan University
Muckle	Gina		Université Laval
Mudra	Susanne		Department of Child and Adolescent Psychiatry, Psychotherapy and Psychosomatics, University Medical Center Hamburg-Eppendorf, Germany
Naerde	Ane		Norwegian Center for Child Behavioral Development
Nakagawa	Astuko		Nagoya City University
Nakano	Shigeru		Sapporo International University
Nakazawa	Jun		Chiba University Emeritus, Uekusa Gakuen University
Nasvytienė	Dalia		Vilnius University
Nava	Elena		University of Milano-Bicocca
Nazzari	Sarah		Brain and Behavioral Sciences Department, University of Pavia, Pavia, Italy
Neppl	Tricia	K.	Iowa State University
Neuenschwander	Regula		University of Bern
Nixon	Elizabeth		Trinity College Dublin
Nolvi	Saara		University of Turku
Nomura	Yoko		City University of New York at Queens College
Nouri	Ali		Malayer University
Ogelman	Hulya	Gulay	Pamukkale University
Ogura	Tamiko		Osaka University of Comprehensive Children Education
Ola	Bolanle	Adeyemi	Department of Behavioral Medicine, College of Medicine, Lagos State University
Olazabal	Daniel		Departamento de Fisiología, Facultad de Medicina, Universidad de la República Uruguay

Last Name	First Name	Middle Name	Institution
Oldehinkel	Albertine	J.	University of Groningen, University Medical Center Groningen
Ong	Ken	K.	University of Cambridge
Oort	Frans		University of Amsterdam
Oosterman	Mirjam		Clinical Child and Family Studies and Amsterdam Public Health research institute, Vrije Universiteit Amsterdam
Oppenheim	David		University of Haifa
Orekhova	Liliia		V.I. Vernadsky Crimean Federal University
Orioli	Guilia		Università degli Studi di Padova
Otte	Renée		Philips Mother & Child Care
Overbeek	Geertjan		University of Amsterdam
Özmert	Elif	N.	Hacettepe University Faculty of Medicine, Department of Pediatrics
Pace	Ugo		Kore University of Enna
Palafoll	Anna	Valenzuela	Barcelona University
Pallini	Susanna		University of Rome Tre
Panescu	Oana		Spiru Haret University
Paquette	Daniel		University of Montreal
Park	Curie		Yonsei Center for Psychological Health
Park	Nayoung		Kyungpook National University
Park	Seong-Yeon		Ewha Womans University
Park	Young-shin		Kyungpook National University
Parkin	Patricia	C.	The Hospital for Sick Children, Dalla Lana School of Public Health, University of Toronto
Pascual-Sagastizabal	Eider		University of the Basque Country
Passini	Christina		University of Geneva
Pavlenko	Vladmir		V.I. Vernadsky Crimean Federal University
Paz	Yael		The Hebrew University of Jerusalem
Peled	Maayan		Haifa University
Peltola	Mikko		Tampere University
Pérez-Edgar	Koraly		The Pennsylvania State University
Pfitzer	Constanze		Department of Congenital Heart Disease—Pediatric Cardiology, German Heart Center Berlin; Berlin Institute of Health (BIH); DZHK (German Center for Cardiovascular Research), partner site Berlin
Pihet	Sandrine		University of Applied Sciences and Arts Western Switzerland
Porges	Stephen		Indiana University
Potharst	Eva		UvA minds, academic outpatient (child and adolescent) treatment center of the University of Amsterdam
Potmesil	Milon		Palacky University Olomouc
Potmesilova	Petra		Palacky University Olomouc
Powell	Stephanie		University of Sheffield
Putnam	Samuel	P.	Bowdoin College
Putnick	Diane		Eunice Kennedy Shriver National Institute of Child Health and Human Development
Quadrelli	Ermanno		University of Milano-Bicocca
Quigley	Jean		Trinity College Dublin
Raikes	Helen		University of Nebraska-Lincoln
Raikkonen	Katri		University of Helsinki

Last Name	First Name	Middle Name	Institution
Rastegar	Pedram	J.	University of Alabama at Birmingham
Rayson	Holly		University of Liverpool and University of Reading
Reynolds	Esther		University of Tennessee-Knoxville
Reynolds	Gemma		Middlesex University
Rodríguez	María	Diana Ruvalcaba	Universidad Autónoma de Nuevo León, Facultad de Enfermería
Rodríguez-Bailón	Rosa		Universidad de Granada
Roebers	Claudia	M.	University of Bern
Roeyers	Herbert		Ghent University - Department of Experimental-Clinical and Health Psychology
Rogers	Samantha	L.	University of Hertfordshire
Rojmahamongkol	Pat		Department of Pediatrics, Siriraj Hospital, Mahidol University
Roskam	Isabelle		Psychological Sciences Research Institute of the University of Louvain
Rossignol	Ana	Sancho	Service de Psychiatrie de l'enfant et de l'Adolescent
Ross-Sheehy	Shannon		University of Tennessee-Knoxville
Rothbart	Mary	K.	University of Oregon
Rudasill	Kathleen	M.	Virginia Commonwealth University
Rueda	M. Rosario		Universidad de Granada
Ruetti	Eliana		Unidad de Neurobiología Aplicada, CEMIC-CONICET
Sacchi	Chiara		University of Padova
Šakan	Dušana		Faculty of Legal and Business Studies, Novi Sad Serbia
Salinas-Quiroz	Fernando		Eliot-Pearson Department of Child Study and Human Development, Tufts University
Salman-Engin	Selin		Bilkent University
Salmivalli	Christina		University of Turku
Salvadori	Eliala		University of Amsterdam
Sándor	Piroska		Semmelweis University
Sandseter	Ellen	Beate Hansen	Queen Maud University College
Sari	Burcu	Akin	Başkent University Medical Faculty, Child and Adolescent Psychiatry Department
Scher	Anat		Haifa University
Scheuring	Noemi		Heim Pal National Paediatric Institute
Schmitt	Katharina		Department of Congenital Heart Disease—Pediatric Cardiology, German Heart Center Berlin; DZHK (German Center for Cardiovascular Research), partner site Berlin
Schuengel	Carlo		Clinical Child and Family Studies and Amsterdam Public Health research institute, Vrije Universiteit Amsterdam
Schults	Astra		University of Tartu
Schwebel	David	C.	University of Alabama at Birmingham
Scott	Shawna	A.	University of Windsor
Sebastian-Galles	Nuria		Universitat Pompeu Fabra
Segretin	Maria	Soledad	Unidad de Neurobiología Aplicada, CEMIC-CONICET
Séguin	Jean	R.	Université de Montréal
Sehic	Ela		Washington State University
Sette	Stefania		Sapienza University of Rome
Shai	Dana		The Academic College Tel Aviv - Yaffo
Shakeel	Aqsa		Fatima Jinnah Women University, Rawalpindi

Last Name	First Name	Middle Name	Institution
Shamama-tus-Sabah	Syeda		Govt. Graduate College, Asghar Mall, Rawalpindi, Pakistan
Shanan	Noga		Department of Education, Ben-Gurion University
Simonelli	Alessandra		University of Padova
Siraj	Attiya		Quaid-i-Azam University
Skalkidou	Alkistis		Uppsala University
Slagt	Meike		Utrecht University
Slobodskaya	Helena		Research Institute of Physiology and Basic Medicine
Smojver-Ažić	Sanja		University of Rijeka
Sodian	Beate		Ludwig-Maximilians-University
Solís-Cámara	Pedro		University of Guadalajara
Spencer	Debra		University of Cambridge
Stanković	Milica		Faculty of Legal and Business Studies, Novi Sad Serbia
Stępień-Nycz	Małgorzata		Jagiellonian University
Stoilova	Antoaneta		New Bulgarian University, Sofia
Stropnik	Staša		University Children's Hospital Ljubljana University Medical Centre
Suarez	Sonia		University of Nebraska-Lincoln
Sukigara	Masune		Nagoya City University
Sullivan	Margaret		Rutgers University, Department of Biomedical and Health Sciences
Szabo	Laszlo		Heim Pal National Paediatric Institute
Szakadát	Sára		SOS Children's Villages
Täht-Kriisa	Eha		University of Tartu
Takács	Lea		Charles University
Tangonan	Dianne	Rose	Philippine Normal University
Tardif	Twila		University of Michigan
Tatai	Csilla		University of Debrecen
Thijs	Carel		Maastricht University Medical Centre+, Care and Public Health Research Centre (CAPHRI), dept of Epidemiology; KOALA Birth Cohort Study, The Netherlands
Thomsen	Tamara		University of Hildesheim
Tibu	Florin		Institute of Child Development
Tóth	Ildikó		Institute of Cognitive Neuroscience and Psychology, HUN-REN Research Centre for Natural Sciences, Budapest
Tough	Suzanne		University of Calgary
Tsiara	Evaggelia		Department of Early Childhood Education, University of Thessaly, Argonafton & Filellinon
Tuovinen	Soile		University of Helsinki
Turati	Chiara		University of Milano-Bicocca
Tutnjević	Slavica		University of Banja Luka, Faculty of Philosophy
Ucus	Sukran		Ahi Evran University, Turkiye
Uka	Fitim		University of Prishtina "Hasan Prishtina" and Psycho-Social and Medical Research Centre
Urbain-Gauthier	Nadine		University Paris
Utsumi	Shoka		Ochanomizu University
Vacaru	Stefania		Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, The Netherlands
Væver	Mette	S.	Center for Early Intervention and Family Studies, Department of Psychology, University of Copenhagen
van Aar	Jolien		University of Amsterdam

Last Name	First Name	Middle Name	Institution
van Aken	Marcel	A.G.	Utrecht University
van Baar	Anneloes	L.	Utrecht University
van den Bergh	Bea	R. H.	KULeuven - University of Leuven
van der Borgh-Sleddens	Ester		Mondriaan Mental Health Center; KOALA Birth Cohort Study, The Netherlands.
van der Poel	Susanne		University of Amsterdam
Van Egeren	Laurie		University of Minnesota
van Liempt	Imre		University of Amsterdam
Vecchio	Giovanni	Maria	University of Roma Tre
Vergara	Ana	I.	University of the Basque Country
Vergara	Vianey	Méndez	National Autonomous University of Mexico
Verhage	Marije	L.	Clinical Child and Family Studies and Amsterdam Public Health research institute, Vrije Universiteit Amsterdam
Verhoeven	Marjolein		Utrecht University
Vierikko	Elina		Tampere University
Virtala	Paula		Centre of Excellence in Music, Mind, Body and Brain Department of Psychology and Logopedics Faculty of Medicine University of Helsinki
Vogel	Franziska		Otto-Friedrich-Universität Bamberg
Wang	Pei-Ling		University of Taipei
Wang	Zhengyan		Capital Normal University
Warreyn	Petra		Ghent University - Department of Experimental Clinical and Health Psychology
Weeland	Joyce		Youth and Family, Erasmus School of Social and Behavioural Sciences, Erasmus University Rotterdam
Wells	Jonathan		UCL Great Ormond Street Institute of Child Health, London
Xing	Xiaopei		Capital Normal University
Yalçınkaya-Alkar	Özden		Ankara Yıldırım Beyazıt University
Yang	Pei-Jung		National Chengchi University
Yi	Hee Yeon		Sesalmaul SSK Research Center, Gachon University
Yirmiya	Nurit		The Hebrew University of Jerusalem
Yoleri	Sibel		İzmir Democracy University
Yurkowski	Kim		University of Ottawa, School of Psychology
Yuval-Adler	Shira		University of Haifa
Zaharia	Corina		Spiru Haret University
Zande	Diāna		University of Latvia
Zappulla	Carla		University of Palermo
Zdebik	Magdalena	A.	Université du Québec en Outaouais
Zhang	Guangzhen		Southeast University
Zubizarreta	Anik		University of Deusto
Zweig	Aaron		Bowdoin College
Zylicz	Piotr	Olaf	Warsaw School of Social Psychology