Article

Leveraging Python to Process Cross-Cultural Temperament Interviews: A Novel Platform for Text Analysis

Journal of Cross-Cultural Psychology
2020, Vol. 51(2) 168–181

DOI: 10.1177/0022022120906478
journals.sagepub.com/home/jcc
SAGE

Joshua J. Underwood[1], Cornelia Kirchhoff[1], Haven Warwick[1], and Maria A. Gartstein[1]

Abstract

During childhood, parents represent the most commonly used source of their child's temperament information and, typically, do so by responding to questionnaires. Despite their wide-ranging applications, interviews present notorious data reduction challenges, as quantification of narratives has proven to be a labor-intensive process. However, for the purposes of this study, the labor-intensive nature may have conferred distinct advantages. The present study represents a demonstration project aimed at leveraging emerging technologies for this purpose. Specifically, we used Python natural language processing capabilities to analyze semistructured temperament interviews conducted with U.S. and German mothers of toddlers, expecting to identify differences between these two samples in the frequency of words used to describe individual differences, along with some similarities. Two different word lists were used: (a) a set of German personality words and (b) temperament-related words extracted from the Early Childhood Behavior Questionnaire (ECBQ). Analyses using the German trait word demonstrated that mothers from Germany described their toddlers as significantly more "cheerful" and "careful" compared with U.S. caregivers. According to U.S. mothers, their children were more "independent," "emotional," and "timid." For the ECBQ analysis, German mothers described their children as "calm" and "careful" more often than U.S. mothers. U.S. mothers, however, referred to their children as "upset," "happy," and "frustrated" more frequently than German caregivers. The Python code developed herein illustrates this software as a viable research tool for cross-cultural investigations.

Keywords

developmental: child/adolescent, methodology, cultural psychology

Several temperament theories have been proposed, and currently, Rothbart's psychobiological model represents the most widely used conceptual framework. This model defines temperament as encompassing individual differences in emotional, motor, and attentional reactivity, as well as self-regulation, linked with underlying neurobehavioral systems subject to contextual influences

[1]Washington State University, Pullman, USA

Corresponding Author:
Joshua J. Underwood, Washington State University, Pullman, WA, USA.
Email: joshua.underwood@wsu.edu

(Rothbart & Derryberry, 1981). Reactivity domains are generally operationalized in terms of response time, intensity, and duration, whereas self-regulation refers to the modulation of reactivity, including withdrawal, inhibition, and executive attention. Reactivity and regulation are viewed as products of biological contributions, maturation, and experience (Gartstein et al., 2016). Although temperament and personality represent distinct domains of individual differences, there is a substantial overlap as well as continuity (Rothbart & Ahadi, 1994). That is, reactivity and regulation observable in infancy and the toddler period translate into later personality characteristics (Gartstein et al., 2016).

Rothbart's (2012) psychobiological model also speaks to the structure of temperament, with higher order/overarching constructs that parallel three to five factors typically derived in personality research. Negative emotionality and positive emotionality/surgency components are relatively consistent across childhood; however, considerable changes occur in the self-regulation domain. In infancy, regulatory capacity/orienting has been defined by attributes that involve parental efforts (e.g., soothability), whereas effortful control, observable starting with the toddler period, reflects increasingly voluntary control of attention afforded by the maturation of the child's executive attention network (Posner et al., 2012). Importantly, these overarching factors comprised fine-grained temperament dimensions, important in their own right (e.g., independent links to later outcomes noted for fear and anger/frustration; Lengua, 2002; Muris, 2006), with developmental transitions on both levels reported in early childhood (Gartstein & Hancock, 2019).

Although a variety of sources can be used to derive information regarding temperament in early childhood, including laboratory observations and psychophysiological markers (e.g., cortisol levels, electroencephalogram recording), parent report was the first measurement approach utilized in this area of scientific inquiry and continues to be the most prominent in literature. In part, this has to do with convenience of being able to ask parents questions, and parents are also in the position to observe their young children in a variety of contexts for considerable amounts of time. As such, parents are able to provide information not accessible through other means of measurement. Parent-reported information concerning child temperament was initially gleaned via the interview format in the ground-breaking New York Longitudinal Study (NYLS; Thomas et al., 1963), with questions subsequently transformed into survey format (e.g., Infant Temperament Questionnaire; Carey & McDevitt, 1978), largely due to the ease of administration and perhaps even more importantly the ease of processing the resulting data. The Thomas et al. (1963) temperament interview took 1 to 2 hr to administer in infancy and 2 to 4 hr in the second year of life, and extensive coding training was required to quantify the transcript into the nine NYLS dimensions of temperament (activity level, approach/withdrawal, adaptability, mood, threshold, intensity, distractibility, rhythmicity, and attention span/persistence). Additional temperament interview protocols, such as the Temperament Characteristics Schedule (Graham et al., 1973) and the Temperament Characteristics Interview (Dunn & Kendrick, 1980), are notable additions to temperament measurement but have not enjoyed widespread use.

Although the interview approach has not been favored in temperament research, it offers several advantages relative to the questionnaire format which should not be overlooked. That is, unlike multiple-choice surveys, the in-person open-ended interview enables flexibility and provides the opportunity to follow up, asking for clarification and additional information, potentially serving as a richer source of temperament data. Questionnaires can produce evidence of patterns among large populations, yet qualitative interview data often deliver more in-depth insights into participant attitudes, thoughts, and actions (Kendall, 2008). In addition, because the interview does not rely on a multiple-choice response format, it is not subject to the same response biases which can limit utility of questionnaires, for example, when individuals tend to acquiesce or stick to mid-range response options. The latter are especially problematic for cross-cultural investigations because these response tendencies can systematically differ across cultures, presenting as true differences in child temperament (Ng & Lee, 2015). Collecting information via parental

interviews regarding child individual differences in reactivity and regulation can illustrate parental ethnotheories and socialization goals, expected to shape parental narratives (Harkness & Super, 2006). That is, parental values and expectations, influenced by cultural orientation, are likely to manifest themselves in the way parents describe their child and in their support for or limits imposed on temperament displays.

The latter is particularly important because of established links between parental ethnotheories, socialization goals, and child social-emotional development (Gartstein & Putnam, 2018). Temperament development is open to experience: Thus, culturally influenced parenting plays a role in shaping this process. Distinct cultural values and patterns of child rearing tend to be relatively stable (Kohnstamm, 1989). Parents within cultural groups are generally motivated to reproduce temperament characteristics in their offspring that are consistent with their values, and those of the cultural group. Thus, individual children are socialized into phenotypical presentations that are desirable, appropriate, or at least tolerable within cultural norms (Kohnstamm, 1989). Insofar as parents in different cultural groups vary in their child-rearing attitudes and behaviors, "culturally preferred phenotypes" for the offspring result. A sizable literature addresses variability in child social-emotional outcomes among vastly different cultures, for example, comparing Western and non-Western societies such as the United States and Japan, that also differ considerably on individualism–collectivism. There has also been an increased interest in comparing more similar Western societies such as the United States and the Netherlands (e.g., Sung et al., 2015). The Sung et al. (2015) study, and subsequently Desmarais et al. (2019), established a number of differences between the United States and the Netherlands, largely culturally and economically similar Western nations, relying on maternal and paternal reports, respectively. Specifically, these differences favored Dutch infants in terms of displaying less negative emotionality and more advanced regulatory capacity. Importantly, findings from large-scale investigations (Gartstein & Putnam, 2018; Putnam & Gartstein, 2017; Super et al., 2008) showed that samples from Northern European countries (e.g., Finland, the Netherlands) demonstrated some of the lowest levels of negative emotionality and relatively greater regulatory capacity.

The vast majority of existing studies relies solely on the questionnaire format, notable exceptions not withstanding (e.g., Super et al., 1996). As already mentioned, interviews provide rich information concerning child reactivity and regulation and are particularly applicable in cross-cultural research aiming to discern differences and communalities across cultures (Super & Gartstein, 2010). The present study addresses this gap in research, examining temperament in U.S. and German toddlers via an interview conducted with mothers and comparing the frequency of temperament-related words between these samples, controlling for the overall number of words in the caregiver narratives. Moreover, this study provides a valuable demonstration of a novel methodology, which could prove to be an important resource for temperament research and cross-cultural investigations more broadly.

This project was designed to examine aspects of temperament that differentiate children in Germany and the United States. Although many similarities were expected, it was hypothesized that significant differences in temperament descriptions would be evident. That is, some differentiation in temperament profiles based on caregivers' descriptions was expected in light of prior research. Despite the fact that Germany and the United States share similarities in their economic and political structure, cultural orientation variability between Germany and the United States has been noted based on dimensions identified by Hofstede (1980, 2011). These dimensions comprised individualism/collectivism, masculinity/femininity, power distance, uncertainty avoidance, long-term orientation, and indulgence/restraint, examined across more than 50 different countries, including Germany and the United States. Individualistic values prioritize personal achievement, success, and self-reliance, rather than group achievement, emphasized in more collectivistic settings. Masculinity values place an emphasis on assertiveness, performance, and success as opposed to quality of life and personal relationships, which are viewed as more

Feminine. Power distance refers to the degree that less powerful people in a society accept inequality in power and view it to be normal. Uncertainty avoidance scores are related to the extent of nervousness/discomfort felt when presented with an ambiguous situation. Long/short-term orientation is a measure of what the culture prioritizes when it comes to goals, whether the focus is on efforts toward future rewards or immediate gratification. The dimension of indulgence versus restraint involves the extent to which individuals are expected to restrain themselves from seeking gratification. Although cultural orientation scores speak to culture-level effects, in our work, these have been consistently associated with child temperament and parenting variables (Gartstein & Putnam, 2018; Putnam & Gartstein, 2017). Comparisons of the United States and Germany using Hofstede's model indicate that the United States scores higher on power distance, indulgence, and individualism; conversely, ratings for Germany are higher for masculinity, uncertainty avoidance, and long-term orientation ("Dimension Data Matrix," n.d.; Hofstede & Minkov, 2010; Taras et al., 2010).

In our prior work, we noted more proximal differences in elements of U.S. and German children's developmental niche, specifically aspects of toddlers' daily routine (Kirchhoff et al., 2014a). For example, parents from the United States appeared more invested in structured play (i.e., play with a purpose), playing with their children more overall, and purchasing more toys. In terms of sleep, U.S. parents had their toddlers nap more than German parents, and unlike German children, U.S. toddlers were frequently allowed to sleep until they woke on their own in the morning. U.S. parents utilized a larger variety of calming strategies, including cuddling, singing, holding, reading, talking softly, special play, sitting with, or offering a special item (e.g., food or blanket). However, German parents were more likely to involve their child in housework activities. It should be noted these elements of the daily routine were more strongly associated with toddler temperament than parental psychology (ethnotheories and socialization goals; Gartstein & Putnam, 2018).

We hypothesized a higher frequency of words associated with negative affectivity and its components (i.e., discomfort, fear, frustration, motor activation, perceptual sensitivity, sadness, shyness, and soothability), as well as fewer terms reflecting effortful control and related traits (i.e., attention focusing, attention shifting, cuddliness, inhibitory control, and low-intensity pleasure) in the descriptions provided by the U.S. sample compared with German caregivers. Prior research has shown that children from Northern European cultures presented with lower distress proneness (i.e., lower in negative affectivity) and were better regulated (i.e., higher in effortful control) relative to an array of cultures, including the United States. (Gartstein & Putnam, 2018; Putnam & Gartstein, 2017; Sung et al., 2015; Super et al., 2008). For example, we anticipated higher frequency of temperament words such as "quiet" and "distressed," associated with negative affectivity, in the descriptions of U.S. children in comparison with their German counterparts. Similarly, fewer instances of temperament words associated with effortful control, such as "careful," were expected in maternal narratives concerning U.S. toddler temperament, relative to the German sample.

Method

Participants

Temperament interview data, along with additional information beyond the scope of this study, were obtained from U.S. ($N = 46$) and German ($N = 49$) samples (see Table 1 for demographic information). The U.S. sample was recruited in Pullman, Washington and Moscow, Idaho inland northwest region. Mothers of 18- to 36-month-old children were enrolled in the study, using a variety of recruitment sources. Interested mothers were invited to take part in this research through local Saturday market outreach, preschools, and daycare programs, as well as Washington State University (WSU) announcement emails. The German sample was recruited in several cities in Germany (Munich, Stuttgart, Düsseldorf, Paderborn, and Muenster), where interested mothers were asked to volunteer for the study at their preschool. All potential participants were

Table 1. Sample Demographics: Statistical Comparisons Performed Using Independent Samples *T* Tests.

Demographics	United States, *M* (*SD*)	German, *M* (*SD*)
Maternal age (years)	33.96 (5.78)	35.99 (6.59)
Mother's years in school	17.48* (2.15)	16.33* (3.08)
Maternal SES	47.76 (25.46)	58.29 (27.19)
Child age (months)	24.87 (4.61)	27.41 (5.40)

Note. Maternal SES scores assigned corresponding to occupational prestige (Nakao & Treas, 1992).
SES = socioeconomic status.
*$p < .05$.

then contacted by telephone or email by the second author (C.K.) and informed about the study. Only mothers who represented the third generation or higher of their families residing in the United States and Germany were recruited. Both samples comprised primarily highly educated caregivers, meeting criteria for middle to upper middle class. In addition, both samples presented with a comparable gender distribution and mean age of children. Demographic differences between the mothers were limited to the average number of years spent in school, with German means significantly lower than those of the United States. This finding can be attributed to the difference in education systems between the United States and Germany. That is, Germany has a dual system of entry into a profession through either a university degree or vocational training after completing high school, with the latter affording a comparable socioeconomic standing. All the mothers in both samples were employed at the time of data collection.

Measures

The temperament interview. The temperament interview (Kirchhoff et al., 2014b; Kirchhoff, Gartstein, Iverson, et al., 2013; Table 2) was designed to elicit information concerning mothers' views of their children's reactive and regulatory tendencies, addressing temperament as defined by the psychobiological approach (Rothbart & Derryberry, 1981). Interviews were administered after caregivers responded to a toddler temperament questionnaire—the Early Childhood Behavior Questionnaire (ECBQ; Putnam et al., 2006), which primed them to think about their child's temperament. This procedure was utilized to facilitate the conversation with caregivers about individual differences apparent in their toddlers. Interviews were recorded, transcribed, and subsequently translated from German to English by a native/bilingual speaker (second author), relying on backtranslation and consultation with another native/bilingual speaker regarding consistency of the resulting language in an attempt to ensure accuracy (World Health Organization [WHO], n.d.; Penã, 2007). The format of the interview included initial open-ended questions, asking the parent to simply describe the child's temperament in their own words. Subsequently, six questions were asked to inquire about the most and least favorite child attributes, reasons for these preferences, and parental responses to the manifestations of these traits.

The ECBQ. The ECBQ by Putnam et al. (2006) is a fine-grained assessment tool, based on the psychobiological definition of temperament proposed by Rothbart and Derryberry (1981). The ECBQ is a parent report measure of child temperament characteristics designed for children between 18 and 36 months of age. Factor analytic work with this instrument has led to the identification of a hierarchical structure, with three overarching factors encompassing 18 fine-grained composites or scales, in turn based on 201 items. The first factor labeled as Surgency/Extraversion (SE) consists of five subscales: Impulsivity, Activity Level, High-Intensity Pleasure, Sociability, and Positive Anticipation. Negative affectivity includes eight subscales: Discomfort, Fear, Motor Activation, Sadness, Perceptual Sensitivity, Shyness, Soothability, and Frustration scales.

Table 2. Temperament Interview Questions.

"We asked you to complete our temperament survey and answer very specific questions about your child and now I would like to give you an opportunity to tell us about _____ in your own words. What's she or he like?"
"Which of these characteristics/traits (*use words that the parent uses to label dimensions of temperament and/or their manifestations*) are your favorites? Which do you enjoy the most?"
"What makes these characteristics/traits (*use words that the parent uses to label dimensions of temperament and/or their manifestations*) your favorite over others?"
What do you do in response when she or he behaves in a way that displays _____ [the favorite trait(s)]?
"Do you have concerns about any of _____'s characteristics/traits (*use words that the parent uses to label dimensions of temperament and/or their manifestations*)? Do any of these make it more difficult to get through your daily routine successfully "?
"In what way are these characteristics/traits (*use words that the parent uses to label dimensions of temperament and/or their manifestations*) challenging? What makes them difficult to manage?"
What do you do in response when she or he behaves in a way that displays [the challenging trait(s)]?

Effortful control is composed of the following five subscales: Inhibitory Control, Attention Shifting, Low-Intensity Pleasure, Cuddliness, and Attention Focusing. The items address toddler temperament characteristics on a 7-point Likert-type scales with responses ranging from *never* to *always*. Scales were generally internally consistent ($\alpha > .70$ except for Impulsivity), demonstrating satisfactory cross-rater agreement (r ranging from .33 to .79) and test–retest reliability (r = .32–.79) (Putnam et al., 2006). Satisfactory internal consistency estimates with samples from Russia and Japan were also obtained (Kolmagorova et al., 2008; Nakagawa et al., 2007), and construct/predictive validity demonstrated with respect to Children's Behavior Questionnaire indicators (Putnam et al., 2008). With respect to translation, we followed an approach similar to the one utilized with the temperament interviews. Specifically, the ECBQ items were translated from English into German by a native/bilingual speaker (second author), relying on backtranslation and consultation with two other native/bilingual speakers regarding consistency of the resulting language in an attempt to ensure accuracy (Penã, 2007; WHO, n.d.). The result of backtranslation efforts was reviewed by several temperament experts to ensure that the content was consistent with originally designed items and their intent. The German version of the ECBQ is feely available through the Rothbart instrument dissemination website: https://research.bowdoin.edu/rothbart-temperament-questionnaires/

German trait word list. A list of 100 German personality trait words (Schönbach, 1972) was developed on the basis of Anderson's (1968) inventory of 555 personality-related terms to aid in cross-cultural research involving German samples. This trait list provided the basis for our first set of analyses. Anderson (1968) first administered the list of 555 trait words to 100 college students to rate how likable these were as personality characteristics and also to rate their "meaningfulness"—how well participants understood the meaning of each word. Schönbach (1972) conducted follow-up research with Anderson's (1968) trait list, selecting 100 words deemed most meaningful and translating these into English—the words utilized in this study.

ECBQ word list. For our second set of analyses, temperament terms derived from the ECBQ (Putnam et al., 2006) were utilized. Specifically, ECBQ items were reviewed by experts from the WSU Temperament Laboratory, with key temperament words selected for analyses, if agreed on by the raters.

This content evaluation produced 29 ECBQ temperament terms specifically reflecting toddler temperament. Of those 29 terms, two items were reverse scored. "Comfortable" appeared in items negatively associated with the negative emotionality factor, for example, as a response (i.e., "seem comfortable?") for the following item: "In situations where she or he is meeting new people, how often did your child," thus was indicative of less social reticence and had to be reversed. "Calm" was similarly reverse scored because of negative associations with SE and negative emotionality, such as in response ("remain calm ?") to the following items: "Before an exciting event (such as receiving a new toy), how often did your child," and "When approached by an unfamiliar person in a public place (for example, the grocery store), how often did your child." The developmental aspect is important, given that the Schönbach (1972) list developed for use with German samples was focused on individual differences manifested in adulthood, some of which are not salient in the toddler period, for example, "unselfish," "tactful," and "well-mannered," to name a few.

Analytic strategy. Our approach was to examine mothers' narratives in the German and the U.S. samples, identifying key temperament-related words and subsequently comparing the frequency of their use. The analytic strategy was twofold, insofar as we started with an established list of words associated with individual differences relevant in the German culture, and also included key words from a temperament instrument designed specifically for toddlers—the ECBQ (Putnam et al., 2006). The ECBQ was developed in the United States and subsequently widely utilized in cross-cultural investigations (e.g., Gartstein & Putnam, 2018).

Python object-oriented programming language with dynamic semantics (Python Software Foundation, n.d.) was utilized, relying on the Natural Language Toolkit (NTLK) designed to process language-based content. Python was chosen for this methodological demonstration over other similar programs, such as Linguistic Inquiry and Word Count (LIWC; Kahn et al., 2007; Tausczik & Pennebaker, 2010), due to its substantial and growing user community with multiple resources available for novice analysts, in addition to its ability to interact with current statistical packages, such as R. Python also has the ability to incorporate established text analysis tools such as the WordNet-Affect tool (Acerbi et al., 2013). Python is an open-source analytic platform, made freely available for download. Additional advantages include scalability, with the Python NTLK able to process large amounts of data, extending scripts to additional narratives and bigger samples, and interoperability, handling data that are structured differently (e.g., web-based information) with ease and efficiency. The Python script was written to analyze each participant's interview transcript applying both lists of personality/temperament trait words with output providing relevant terms found in the narrative in a descending order based on frequency of use. The script (Appendix) used in the analyses described above was compiled relying on online resources, such as the In-depth User's Manual and Tutorial (Python Software Foundation, n.d.), made available from the publishers upon downloading software. The scripts for both personality/temperament trait word lists were essentially identical, with the only difference being text in the line "File2," determining which word list to use when analyzing the interview. All language content analyses were performed in English.

Results

First, we relied on the Schönbach's (1972) list of trait words, conducting independent-groups *t* tests to compare German and U.S. samples in terms of their frequency of use. These analyses produced five significant differences of 100 personality words examined (Table 3). Specifically, German parents described their toddlers as significantly more "cheerful" and "careful" compared with U.S. caregivers. According to U.S. mothers, their children were more "independent" and "emotional," as well as "timid."

In an effort to ensure that age-appropriate trait words were considered, independent *t* tests were also conducted with frequency counts derived from the caregiver interviews on the basis of

Table 3. Independent Samples *T* Tests Comparing Frequency of Key German Trait Words (Schönbach, 1972) for U.S. and German Caregivers.

	United States		Germany			
Temperament word	*M*	*SD*	*M*	*SD*	*t* test	Cohens *d*
Cheerful	0.02	0.15	0.76	0.95	5.19**	1.50
Independent	0.70	1.26	0.18	0.44	2.67**	0.56
Friendly	0.04	0.21	0.27	0.86	1.70[a]	0.36
Careful	0.00	0.00	0.16	0.43	2.60**	0.52
Emotional	0.20	0.65	0.00	0.00	2.09*	0.44
Timid	0.08	0.28	0.00	0.00	2.00*	0.41
Angry	0.50	1.23	1.00	1.58	1.71[a]	0.35

[a]Trend.
*$p < .05$. **$p < .01$.

the ECBQ temperament words (Table 4). Again, a number of significant differences were observed; specifically, five of the 29 terms were associated with significant differences. German mothers described their children more often as "calm" and "careful" compared with U.S. mothers, whereas U.S. mothers referred to their children as "upset," "happy," and "frustrated" more frequently. Thus, although effect sizes were generally comparable, comparisons utilizing this more age-appropriate word list resulted in a higher proportion of significant differences than the trait word list derived for use with adults (i.e., Schönbach, 1972).

Discussion

As hypothesized, analyses of caregiver narratives using the Schönbach (1972) individual difference trait words, developed with German samples, as well as the trait list derived from the U.S. toddler temperament questionnaire, the ECBQ, revealed a number of significant differences. Specifically, based on the Schönbach (1972) list, German parents described their toddlers as significantly more "cheerful" and "careful" compared with U.S. caregivers. According to U.S. mothers, their children were more "independent" and "emotional," as well as "timid." Using the word list derived from the ECBQ, U.S. narratives represented their children in terms of trait words linked to negative emotionality: "frustrated" and "upset," more frequently than German caregivers. Another trait word that German narratives used significantly more frequently compared with U.S. descriptions was "calm," indicative of lower reactivity and linked to both SE and negative emotionality. The trait word "comfortable" is a reverse scored item, and as such, lower levels of comfort translate to higher levels of negative emotionality. U.S. narratives also demonstrated a significantly more frequent use of the word "happy," which is associated with SE in the ECBQ. Finally, German narratives referred to children in terms of a trait word associated with effortful control more frequently, describing toddlers as "careful" more often than their U.S. counterparts.

Proportionally, more significant differences emerged in the analyses of ECBQ (Putnam et al., 2006) words (five of 29) versus the Schönbach (1972) word list (five of 100). This disproportionate distribution of significant findings favoring the ECBQ-derived word list is likely a function of the latter individual difference terms being more age appropriate for the sample included in this study. It should be noted that a number of temperament/personality words differentiating U.S. and German temperament narratives are semantically related, such as "timid" and "careful," and consistency in results between synonymous and closely related words supports the qualitative data reduction approach illustrated herein. Not surprisingly, observed differences were circumscribed, with a number of similarities in temperament descriptions emerging for the two samples from Western cultures that share many aspects of their economic and political systems.

Table 4. Independent Samples *T* Tests Comparing Frequency of Key ECBQ-Based Trait Words for U.S. and German Caregivers.

	United States		Germany			
Temperament word	*M*	*SD*	*M*	*SD*	*t* test	Cohens *d*
Calm	0.80	1.24	1.53	1.72	−2.34**	0.48
Upset	1.13	2.39	0.31	0.77	2.29**	0.47
Happy	2.76	4.18	1.35	1.59	2.20**	0.45
Careful	0.00	0.00	0.16	0.43	−2.60**	0.52
Frustrated	0.76	1.30	0.14	0.41	3.16**	0.65
Frightened	0.00	0.00	0.06	0.24	−1.71[a]	0.35
Comfortable (R)	0.22	0.47	0.06	0.32	1.91[a]	0.40
Angry	0.50	1.23	1.00	1.58	1.71[a]	0.35

Note. ECBQ = Early Childhood Behavior Questionnaire; *R* = reverse scored.
[a]Trend.
*$p < .05$. **$p < .01$.

Significant differences in trait word counts that were derived from caregiver narratives based on the German personality word list (Schönbach, 1972) are in part consistent with previously reported cross-cultural comparisons. The observed pattern of results is consistent with the findings reported by Harkness et al. (2000) with respect to U.S. and Dutch samples, wherein U.S. parents described their children in a manner that emphasized individualistic characteristics, whereas Dutch caregivers often focused on their children as being socially connected. More frequent mentions of independence in the U.S. temperament narratives can be thought of as consistent with the United States being higher on individualism compared with Germany (Hofstede, 2001). German caregivers' more frequent descriptions of their children as careful could be reflective of a higher rating for Germany on the dimension of long-term orientation (Hofstede & Minkov, 2010). Although individualism/collectivism differences between Germany and the United States are not so great as variability for some of the cultures previously examined with respect to child temperament (e.g., East/West comparisons; Chen et al., 1998; Krassner et al., 2017), these nonetheless appear consequential in the context of this study and elsewhere (Kirchhoff, Gartstein, Brekke, et al., 2013; Kirchhoff, Gartstein, Iverson, et al., 2013).

Analysis using the ECBQ-derived key words indicated that U.S. caregivers used "upset" in describing their children more often, and German parents described their children as "calm" with greater frequency in the context of the temperament interviews. The present findings are, thus, in part consistent with prior questionnaire-based research, wherein German mothers rated their children significantly higher in soothability (Kirchhoff, Gartstein, Brekke, et al., 2013; Kirchhoff, Gartstein, Iverson, et al., 2013), and prior studies indicate that samples from Northern Europe demonstrate superior regulation relative to other regions (Gartstein & Putnam, 2018; Putnam & Gartstein, 2017). U.S. infants being described as happy more frequently may also be a function of a greater cultural emphasis on individualism, which has been widely linked with extraversion (Hofstede & McCrae, 2004), in turn associated with expression of positive affect.

In addition to being related to broader cultural orientation, as documented in prior research (Gartstein & Putnam, 2018; Hofstede & McCrae, 2004; Putnam & Gartstein, 2017), these differences in temperament descriptions for children in the United States and Germany are also likely a function of variability in the developmental niche, and elements of toddlers' daily routine (Kirchhoff et al., 2014) in particular. For example, more frequent mentions of child independence by mothers from the United States could be function of play and sleep routines, wherein children are expected to engage in structured play, often with an educational purpose which demands some autonomy, and generally sleeping until waking on their own in the morning. As German mothers tended to involve their toddlers in housework activities to a greater extent, such activities likely increased the

importance of being careful and provided more opportunities for mothers to observe their children being careful. U.S. parents' reliance on a larger variety of calming strategies (e.g., cuddling, singing, holding, reading, talking softly, special play, sitting with, offering a special item) could in turn make it less imperative for children to lower their own level of arousal, appearing calm/not upset.

Our findings contribute to the existing literature addressing cross-cultural differences in temperament and also provide an important demonstration of an advanced computational technique, leveraged here to quantify interview data concerning child individual differences. However, several limitations should be noted and require additional research. Results reported herein can be considered generalizable to the communities from which the families were recruited, rather than the United States and Germany as a whole. These data speak to characteristics of children whose families represent primarily middle or upper middle class and, thus, may not be reflective of children living in more economically disadvantaged circumstances. In addition, the sample sizes were relatively small, and future research should recruit larger more economically diverse samples, expanded to other cultures. In addition to limiting generalizability, the relatively small sample size also negatively affected statistical power of between-group comparisons conducted in this study. Post hoc power analyses indicated an average power value of .10 for the 100 German trait word (Schönbach, 1972) comparisons, with analyses of the ECBQ trait list (Putnam et al., 2006) resulting in an average power of .28 across 29 comparisons. In light of this limited power, we chose to forgo adjustments for multiple statistical tests, and it will be important for future research to include larger samples, performing such corrections, as needed. The two techniques for identifying temperament/personality terms used in this study, the ECBQ (Putnam et al., 2006) and the German trait list (Schönbach, 1972), showed limited convergence with respect to words that differentiated the two cultural groups. This pattern of results is not surprising as the Schönbach (1972) word list was developed for research with adult German samples, and the ECBQ words were derived from a questionnaire developmentally appropriate for use with 18- to 36-month-old children. Both approaches are informative in their own right, and the observed results speak to the importance of being mindful of developmental considerations in generating emotion word lists.

It should also be noted that results of this study speak to caregiver perceptions, influenced by parental ethnotheories and socialization goals, which represent important cultural distinctions, but do not necessarily reflect differences in observable emotional expression and behaviors. Thus, future research should consider child behaviors alongside parental narratives, the focus on the present investigation. Direct translations of emotion and temperament words can be problematic, and we have encountered this issue in the several translations undertaken with temperament questionnaires (e.g., Gartstein et al., 2003). Although steps were taken in this study to ensure translation accuracy (i.e., translation and backtranslation by a native/bilingual speaker, consultation with another native/bilingual speaker), future research should seek to incorporate a more extensive translation consultation (e.g., the committee method described by Epstein et al., 2015). Finally, participants in this study were primed in regard to their child's temperament by completing the ECBQ (Putnam et al., 2008) prior to the interview. Although both samples were subject to identical procedures, thus primed in the exact same manner, with the latter not contributing to systematic differences between cultural groups, future research should consider reversing the procedure, administering the temperament interview first.

In summary, the present study addresses an important gap in research, utilizing interview methodology for the purposes of a cross-cultural comparison of U.S. and German toddlers. Importantly, development of the Python code utilized for trait words frequency analyses bodes many potential future applications. This study represents a methodological demonstration, potentially leading to more extensive use and analysis of interview data in addition to questionnaires, which have dominated cross-cultural literature concerning individual differences. Large quantities of interview data are difficult to obtain for a variety of reasons, including attrition, interviewee cooperation, and data management issues, such as transcription (Ryen, 2001). However, it is the time and effort-intensive processing and quantification of interview narratives that represent key barriers to more widespread

use. The Python code presented herein demonstrates an accessible approach enabling researchers to overcome these notable difficulties. For instance, larger scale cross-cultural collaborative studies can make use of this open-source readily available program to aid narrative processing/quantification in multiple languages, circumventing the need for translation by incorporating relevant word lists in each of the languages. To aid further research using this method, future studies should also focus on generating relevant terms, as word lists are critical to this analytic approach.

Appendix

The Python Code Developed for Examination of Cross-Cultural Interview Data.

```
File1 = open("D54.txt", "r")
File2 = open("ECBQ Word Set.txt", "r")
Filewords = {}
badSymbols = (".", ",", "!", "?", "'", '"', '“', "participant", "they")
totalCount = 0
over3Count = 0

applicableWords = []
for line in File2:
   applicableWords.append((line.lower()).strip())

for line in File1:
   if not line.startswith("Dr. G"):
      for word in line.split( ):
         modWord = (word.lower())
         for symbol in badSymbols:
            modWord = modWord.replace(symbol, "")
         if len(modWord) > 3:
            if modWord in Filewords:
               Filewords[modWord] += 1
               totalCount += 1
               over3Count += 1
            else:
               Filewords[modWord] = 1
               totalCount += 1
               over3Count += 1
         else:
            totalCount += 1

newList = []
for word in Filewords:
   if word in applicableWords:
      if len(newList) == 0 or Filewords[word] == 1:
         newList.append(word)
      else:
         for item in newList:
            if not word in newList and Filewords[word] >= Filewords[item]:
               newList.insert(newList.index(item), word)

for value in newList:
   print (value + ": " + str(Filewords[value]))

print()
print("Total number of words participant said: " + str(totalCount))
File1.close();
File2.close();
```

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iD


Joshua J. Underwood https://orcid.org/0000-0002-1941-3097

References

Acerbi, A., Lampos, V., Garnett, P., & Bentley, R. A. (2013). The expression of emotions in 20th century books. *PLOS ONE*, *8*(3), e59030. https://doi.org/10.1371/journal.pone.0059030

Anderson, N. (1968). Likableness ratings of 555 personality-trait words. *Journal of Personality and Social Psychology*, *9*(3), 272–279.

Carey, W. B., & McDevitt, S. C. (1978). Revision of the Infant Temperament Questionnaire. *Pediatrics*, *61*(5), 735–739.

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*(4), 677–686. https://doi.org/10.1037/0012-1649.34.4.677

Desmarais, E., Majdandžić, M., Gartstein, M. A., Bridgett, D. J., & French, B. F. (2019). Cross-cultural differences in temperament: Comparing paternal ratings of US and Dutch infants. *European Journal of Developmental Psychology*, *16*(2), 137–151. https://doi.org/10.1080/17405629.2017.1356713

Dimension data matrix. (n.d.). https://geerthofstede.com/research-and-vsm/dimension-data-matrix/

Dunn, J., & Kendrick, C. (1980). The arrival of sibling: Changes in patterns of interaction between mother and first-born child. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *21*(2), 119–132.

Epstein, J., Osborne, R. H., Elsworth, G. R., Beaton, D. E., & Guillemin, F. (2015). Cross-cultural adaptation of the Health Education Impact Questionnaire: Experimental study showed expert committee, not back-translation, added value. *Journal of Clinical Epidemiology*, *68*, 360–369.

Gartstein, M. A., Carranza, J., González-Salinas, C., Ato, E., Galián, M., Erickson, N., & Potapova, N. (2016). Cross-cultural comparisons of infant fear: A multi-method study in Spain and the United States. *Journal of Cross-Cultural Psychology*, *47*(9), 1178–1193.

Gartstein, M. A., & Putnam, P. (2018). *Toddlers, parents, and culture: Findings from the joint effort toddler temperament consortium*. Routledge.

Gartstein, M. A., Slobodskaya, H. R., & Kinsht, I. A. (2003). Cross-cultural differences in the first year of life: United States of America (U.S.) and Russian. *International Journal of Behavioral Development*, *27*, 316–328.

Gartstein, M. A., & Hancock, G. R. (2019). Temperamental growth in infancy: Demographic, maternal symptom, and stress contributions to overarching and fine-grained dimensions. *Merrill-Palmer Quarterly*, *65*(2), 121–157. Article 1. https://doi.org/10.13110/merrpalmquar1982.65.2.0121

Graham, P., Rutter, M., & George, S. (1973). Temperamental characteristics as predictors of behavior disorders in children. *American Journal of Orthopsychiatry*, *43*(3), 328–339.

Harkness, S., & Super, C. M. (2006). Themes and variations: Parental ethnotheories in Western Cultures. In K. H. Rubin & O. B. Chung (Eds.), *Parenting beliefs, behaviors, and parent-child relations: A cross-cultural perspective* (pp. 61–79). Psychology Press.

Harkness, S., Super, C. M., & Tijen, N. (2000). Individualism and the "western mind" reconsidered: American and Dutch parents' ethnotheories of the child. In S. Harkness, C. Raeff, & C. M. Super (Eds.), *New directions for child and adolescent development: Variability in the social construction of the child* (Vol. 87, pp. 23–39). Jossey-Bass.

Hofstede, G. (1980). *Culture's consequences: International differences in work-related values*. Beverly Hills, CA: Sage.

Hofstede, G. (2001). *Culture's consequences: comparing values, behaviors, institutions, and organizations across nations* (2nd ed.). Thousand Oaks, CA: Sage.

Hofstede, G. (2011). Dimensionalizing cultures: The Hofstede model in context. *Online Readings in Psychology and Culture*, *2*(1), 8. https://doi.org/10.9707/2307-0919.1014

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

Hofstede, G., & Minkov, M. (2010). Long- versus short-term orientation: New perspectives. *Asia Pacific Business Review*, *16*(4), 493–504. https://doi.org/10.1080/13602381003637609

Kahn, J. H., Tobin, R. M., Massey, A. E., & Anderson, J. A. (2007). Measuring emotional expression with the Linguistic Inquiry and Word Count. *The American Journal of Psychology*, *120*(2), 263–286.

Kendall, L. (2008). The conduct of qualitative interview: Research questions, methodological issues, and researching online. In J. Coiro, M. Knobel, C. Lankshear, & D. Leu (Eds.), *Handbook of research on new literacies* (pp. 133–149). Lawrence Erlbaum.

Kirchhoff, C. M., Gartstein, M. A., Brekke, A., & Cress, E. (2013, April). *Temperament attributes, parent-child interaction factors, and behavior problems in the U.S. and Germany* [Paper presentation]. Biennial Convention of the Society for Research in Child Development, Seattle, WA, United States.

Kirchhoff, C. M., Gartstein, M. A., Iverson, S. L., & Johnson, S. (2014b, July). *Is toddler temperament influenced by the context of parent-child interactions: Differences between Germany and the U.S*. [Paper presentation]. Biennial Convention of the International Society on Infant Studies, Berlin, Germany.

Kirchhoff, C. M., Gartstein, M. A., Iverson, S. L., Reed-Smith, H., Brekke, A., & Cress, E. (2013, January). *Temperament and behavioral difficulties across cultures: A comparison between German and U.S. toddlers* [Paper presentation]. Occasional Temperament Conference, Salt Lake City, UT, United States.

Kirchhoff, C. M., Krassner, A. M., Putnam, S., & Gartstein, M. A. (2014a, July). *Exploring the developmental niche: Differences between U.S. and Germany* [Paper presentation]. Biennial Convention of the International Society on Infant Studies, Berlin, Germany.

Kohnstamm, G. A. (1989). *Temperament in childhood: Cross-cultural and sex differences*. John Wiley.

Kolmagorova, A. V., Slobodskaya, H. R., & Gartstein, M. A. (2008). A Russian adaptation of the Early Childhood Behavior Questionnaire (ECBQ) for measuring temperament. *Psikhologicheskii Zhurnal*, *29*, 121–136.

Krassner, A., Gartstein, M. A., Park, C., Wojciech, D., Lecannelier, F., & Putnam, S. (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.

Lengua, L. J. (2002). The contribution of emotionality and self-regulation to the understanding of children's response to multiple risk. *Child Development*, *73*, 144–161. https://doi.org/10.1111/1467-8624.00397

Muris, P. (2006). The pathogenesis of childhood anxiety disorders: Considerations from a developmental psychopathology perspective. *International Journal of Behavioral Development*, *30*, 5–11. https://doi.org/10.1177/0165025406059967

Nakagawa, A., Sukigara, M., & Mizuno, R. (2007, April). *Cultural effects reflected in the Early Childhood Behavior Questionnaire for Japanese toddlers: Psychometrics and factor structure* [Paper presentation]. Society for Research in Child Development Biennial Meeting, Boston, MA, United States.

Nakao, K., & Treas, J. (1992). *The 1989 socioeconomic index of occupations: Construction from the 1989 Occupational Prestige Scores* (General Social Survey Methodological Reports No. 74). National Opinion Research Center.

Ng, S., & Lee, A. Y. (2015). *Frontiers of culture and psychology: Handbook of culture and consumer behavior*. Oxford University Press.

Penã, E. D. (2007). Lost in translation: Methodological considerations in cross-cultural research. *Child Development*, *78*(4), 1255–1264.

Posner, M. I., Rothbart, M. K., Sheese, B. E., & Voelker, P. (2012). Control networks and neuromodulators of early development. *Developmental Psychology*, *48*(3), 827–835.

Putnam, S. P., & Gartstein, M. A. (2017). Aggregate temperament scores from multiple countries: Associations with aggregate personality traits, cultural dimensions, and allelic frequency. *Journal of Research in Personality*, *67*, 157–170.

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 & Development*, *29*(3), 386–401.

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*(4), 387–405.

Python Software Foundation. (n.d.). *The python tutorial—Python documentation*. https://docs.python.org/3/tutorial/

Rothbart, M. K., & Ahadi, S. A. (1994). Temperament and the development of personality. *Journal of Abnormal Psychology*, *103*(1), 55–66.

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). Lawrence Erlbaum.

Rothbart, M. (2012). Advances in temperament: History, concepts, and measures. In M. Zentner & R. L. Shiner (Eds.), *Handbook of temperament* (pp. 3–20). The Guilford Press.

Ryen, A. (2001). Cross-cultural interviewing. In J. F. Gubrium & J. A. Holstein (Eds.), *Handbook of interview research* (pp. 334–354). SAGE. https://doi.org/10.4135/9781412973588.n22

Schönbach, P. (1972). Likableness ratings of 100 German personality-trait words corresponding to a subset of Anderson's 555 trait words. *European Journal of Social Psychology*, *2*(3), 327–333.

Sung, J., Beijers, R., Gartstein, M. A., de Weerth, C., & Putnam, S. P. (2015). Exploring temperamental differences in infants from the United States of America (U.S.) and the Netherlands. *European Journal of Developmental Psychology*, *12*(1), 15–28.

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*(1–2), 136–157. https://doi.org/10.3233/DEV-2008-21209

Super, C. M., & Gartstein, M. A. (2010, October). *Preconference on interpretation of cross-cultural data* [Paper presentation]. Occasional Temperament Conference, Brunswick, ME, United States.

Super, C. M., Harkness, S., Van Tijen, N., Van Der Vlugt, E., Fintelman, M., & Dijkstra, J. (1996). The three R's of Dutch childrearing and the socialization of infant arousal. In S. Harkness & C. M. Super (Eds.), *Parents' cultural belief systems: Their origins, expressions, and consequences* (pp. 447–466). Guilford Press.

Taras, V., Kirkman, B. L., & Steel, P. (2010). Examining the impact of culture's consequences: A three-decade, multi-level, meta-analytic review of Hofstede's cultural value dimensions. *Journal of Applied Psychology*, *95*, 405–439.

Tausczik, Y. R., & Pennebaker, J. W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology*, *29*, 24–54. https://doi.org/10.1177/0261927X09351676

Thomas, A., Chess, S., Birch, H. G., Hertzig, M. E., & Kom, S. (1963). *Behavioral individuality in early childhood*. New York University Press.

World Health Organization. (n.d.). *Management of substance abuse project: Process of translation and adaptation of instruments*. http://www.who.int/substance_abuse/research_tools/translation/en/