

Relations between Temperament and Theory of Mind Development in the U.S. and China:
Biological and Behavioral Correlates of Preschoolers' False-Belief Understanding

Jonathan D. Lane, Henry M. Wellman, Sheryl L. Olson, Alison L. Miller
University of Michigan

Li Wang
Peking University

Twila Tardif
University of Michigan

For the published version of this manuscript, please see:

[Lane, J. D., Wellman, H. M., Olson, S. L., Miller, A. L., Wang, L. & Tardif, T. \(2013\). Relations between temperament and theory of mind development in the United States and China: Biological and behavioral correlates of preschoolers' false-belief understanding. *Developmental Psychology*, 49, 825-836. doi: 10.1037/a0028825](#)

This research was supported by grant SES 0527475 from the National Science Foundation to Twila Tardif (PI). We are very grateful to the children, parents, teachers, and preschool administrators who participated, and to the many individuals who gave us invaluable help with data collection and coding: Daniel Kessler, Tassie Hajal, Dana Richter, Liz Kelly, Jesse Wilkinson, Cristin Brophy, Emily Youra, Juli Harrison, Midori Kazama, Helen Y. J. Lee, Akiko Hayashi, Saori Kumasaka, Yukiko Ishii, Chika Nakamura, Sachiko Ohno, Hiromi Ikeda, and Keiko Yamaguchi.

Abstract

The emotional reactivity hypothesis holds that, over the course of phylogeny, the selection of animals with less reactive temperaments supported the development of sophisticated social-cognitive skills in several species, including humans (Hare, 2007). In the ontogenetic human case, an emotional reactivity hypothesis predicts that children with less reactive temperaments will reach certain milestones in theory-of-mind (ToM) development more quickly. We examined relations between temperament and false-belief understanding in 102 preschool-age children from China and the United States. Temperament was measured via parental ratings of behavior as well as with physiological measures of children's reactivity (HPA-axis reactivity gauged via salivary cortisol). In accord with an emotional reactivity hypothesis, children with certain reactive temperaments—specifically, those who were more aggressive and those who were both socially-withdrawn and physiologically-reactive—evidenced poorer social-cognition. However, our findings also force amendment to the ontogenetic emotional reactivity hypothesis. For the majority of children in both countries, physiological reactivity predicted more advanced ToM, perhaps by facilitating social engagement and attention to social stimuli. Moreover, children who were withdrawn from social interaction yet non-reactive, especially Chinese children of this temperament, evidenced advanced ToM. Thus, some forms of social disengagement may foster social-cognitive development in certain socio-cultural contexts.

Key words: Theory-of-mind, temperament, emotional reactivity

Relations between Temperament and Theory of Mind Development in the U.S. and China:**Biological and Behavioral Correlates of Preschoolers' False-Belief Understanding**

Children's appreciation for how others think, feel, and behave—their theory of mind (ToM)—undergoes dramatic development during early childhood (Harris, 2006; Wellman, 1990). Such development depends on the quality of social interaction in which children engage or observe (Astington, 2005; Carpendale & Lewis, 2004). For example, preschoolers with older siblings—siblings who typically possess more sophisticated social-cognitive abilities—exhibit a more advanced ToM themselves (Perner, Ruffman, & Leekham, 1994). Parents' reference to mental and emotional states in their conversations with children also predicts children's more advanced ToM (Astington & Jenkins, 1995; Ruffman, Slade, & Crowe, 2002). However, other people are not the only ones responsible for the social interactions in which children engage—children's own competencies and temperaments are also important although much less studied with regard to ToM development. We focus on the role of childhood temperament in ToM development, building on emerging research that has begun to establish such a link. Importantly, we add physiological measures to the behavioral temperament ratings used in prior research to confirm and clarify predictive links between young children's temperament and their ToM development.

Our study, on social-cognitive development in children, was inspired by the *emotional reactivity hypothesis*, first developed to account for the evolution of some of dogs' human-like social-cognitive skills (Hare & Tomasello, 2005a), and since applied to the phylogeny of social-cognition in apes (e.g., Hare, 2007) and the ontogeny of social-cognition in human children (e.g., Wellman, Lane, Labounty, & Olson, 2011). In short, dogs outperform wild wolves and foxes on various social-communicative tasks (Hare, Brown, Williamson, & Tomasello, 2002; Hare &

Tomasello, 2005a; Kamininski, Call, & Fischer, 2004; Soproni, Miklósi, Topál, & Csányi, 2001). The emotional reactivity hypothesis posits that wild canines that were nonaggressive and less fearful of humans were more welcomed in the company of humans and selectively bred over generations. Because of their comfort around humans, these canines were better suited to attend to and interpret humans' gestures and social cues (Hare & Tomasello, 2005b). Thus, one product of domestication was social-communicative abilities in dogs that resemble some of the early social-cognitive abilities of human infants.

Recent work on chimpanzees and bonobos (our closest evolutionary cousins) lends support to an emotional reactivity hypothesis with respect to primate phylogeny. Although chimpanzees and bonobos are similar in many respects, they demonstrate quite different temperaments, with bonobos characterized as less aggressive, more tolerant, and shyer (de Wall, 2005; Hare, Melis, Woods, Hastings, & Wrangham, 2007; Herrmann, Hare, Call, & Tomasello, 2010). These differences in temperament intriguingly relate to cognitive differences that correspond with the emotional reactivity hypothesis. For example, bonobos outperform chimpanzees on cooperative communicative tasks and social-cognitive tasks, but not on non-social cognitive tasks (Hare et al., 2007; Herrman et al., 2010; Okamoto-Barth, Call, & Tomasello, 2007). Thus, Hare and colleagues (Hare, 2007; Hare et al., 2007) propose that emotional reactivity influenced the evolutionary development of ToM in apes, including humans.

Conceivably, a less reactive temperament might also facilitate the ontogenetic development of social-cognition in human children. This version of an emotional reactivity hypothesis was first examined by Wellman et al. (2011), who reasoned that children with certain temperaments may be more comfortable around and attentive to others, and thus better suited to detect and learn from others' social cues. Several aspects of their longitudinal data are consistent

with such a hypothesis. Specifically, parents' ratings of children's shyness, perceptual sensitivity, and lower levels of physical aggression at age 3-years predicted children's more advanced false-belief understanding two years later. Wellman and colleagues (2011) argued that a nonreactive, socially-observant temperament enhances ToM development during the preschool years. Several other studies report related findings (e.g., Capage & Watson, 2001; Slaughter, Dennis, and Pritchard, 2002). For example, Capage and Watson (2001) also found that preschoolers' physical aggression was negatively related to their ToM.

These results require confirmation and expansion. The prior results all depend on parental or teacher ratings of children's temperament. Such ratings have proven valuable but are also limited, as we discuss shortly. Behavioral ratings of temperament could be complemented by other types of data, particularly physiological measures of temperament. For example, it is known that childhood temperament and behavior (particularly characteristics such as shyness and aggression) vary with activation of the hypothalamic-pituitary-adrenocortical (HPA) axis (Kagan & Snidman, 2004; Lopez, Vazquez, & Olson, 2004; Rothbart & Bates, 2006)—a system that is involved in modulating individuals' physiological and behavioral reactions to stress (Lopez et al., 2004), as well as their energy metabolism (Pervanidou & Chrousos, 2011) and sleep regulation (Clow, Hucklebridge, Stalder, Evans, & Thorn, 2010). In the non-human case, the evolution of temperaments that are more suitable to developing advanced social-cognition has been credited to the “domestication” of the HPA-axis (Belyaev, 1979; Hare, 2007; Trut, Oskina, & Kharlamova, 2009). Cortisol is a product of HPA-axis activation, and individual differences in cortisol serve as reliable indicators of emotion regulation problems and temperament (Gunnar et al., 1997; Gunnar et al., 2003; Lopez et al., 2004). Thus, in the current

study, we use children's salivary cortisol as well as behavioral measures of temperament to predict children's social-cognitive development.

The nature of a relationship between physiological reactivity and social-cognitive development may take several forms. In clear accord with the emotional reactivity hypothesis, greater physiological reactivity could predict poorer social-cognition by limiting quality social interactions. More pronounced HPA-axis reactivity is predictive of peer rejection (Gunnar et al., 2003), perhaps because reactivity is also associated with reactive physical aggression (Lopez-Duran, Olson, Hajal, Felt, & Vazquez, 2009; Murray-Close, Han, Cicchetti, Crick, & Rogosch, 2008) as well as with the expression of more anger in response to disappointment (Donzella, Gunnar, Krueger, & Alwin, 2000). Alternatively, physiological reactivity could positively relate to social-cognition, at least for some children. For example, some reactivity to new social situations is associated with peer acceptance (Gunnar et al., 1997)—thus a certain level of reactivity may facilitate positive social interactions which foster social-cognitive development. Notably, in a rare study assessing relations between physiological reactivity and social-cognition, Lewis and Ramsay (1997) found that cortisol response was positively associated with infants' mirror self-recognition. Further research is nonetheless needed to clarify these various possible relations between reactivity and social-cognitive development.

Beyond assessing direct relations between HPA-axis reactivity and social-cognition, a physiological measure of reactivity has potential advantages when used in conjunction with behavioral ratings: it can help to distinguish between children who may have different temperaments, but who receive similar scores on traditional behavioral temperament ratings. This could be especially informative when considering children's shyness or social withdrawal, because withdrawal has several hypothesized subtypes (e.g., Asendorpf, 1990; Rubin & Mills,

1988), and children of these different subtypes exhibit different social-cognitive abilities. For example, U.S. kindergarteners who Harrist and colleagues (1997) classified as *active-isolate* (those who were more angry and less inhibited, according to teacher ratings) faced greater peer rejection and demonstrated social-cognitive deficits when compared to *passive-anxious* children—those who were more timid and withdrawn, but not reactive or defiant. More recently, Chen, Wang, and Cao (2011) found that Chinese elementary-school children classified as *unsociable* (those who lack interest in others, according to peer ratings) experience poorer psychological adjustment and lower social standing than *shy-sensitive* children, who are interested in others yet timid. Differences in the types of withdrawn children identified by different measures and captured in different samples may account for conflicting prior findings: for example, whereas Wellman and colleagues (2011) found that parent reports of children's shy/withdrawn behavior predicted better ToM, Walker (2005) found a negative correlation between shy/withdrawn behavior and ToM (see also Banerjee & Henderson, 2001). Thus, it is potentially important to differentiate between children who exhibit different forms of social withdrawal. Yet, some of the most widely-used ratings of children's temperament (including those used in the current study) do not differentiate between forms of social withdrawal. For example, items on the *withdrawal* scale of the widely-used Child Behavior Checklist (CBCL, Achenbach, 1992)—our observational measure of children's social withdrawal—could refer to a reactively-withdrawn or to a shy temperament profile (e.g., “Withdrawn, doesn't get involved with others”; “Doesn't answer when people talk to him/her”; “Avoids looking others in the eye”). Using a physiological indicator of temperament along with these behavioral measures will allow us to better understand the relations between children's social withdrawal and ToM development.

Insight into the relations between ToM development and temperament requires generalizing findings to children growing up in different socio-cultural contexts. Because of large differences in their historic, linguistic and cultural backgrounds, comparisons between children raised in China and the U.S. have proven important for establishing similarities and differences in ToM development. This research has demonstrated that early ToM development has a similar trajectory in these and other cultures (Wellman et al., 2001), and false-belief understanding in particular has a very similar developmental trajectory in the U.S. and China (Liu, Wellman, Tardif, & Sabbagh, 2008; Tardif, Wellman, & Cheung, 2004). Yet research on relations between temperament and ToM development has, thus far, only been conducted with Western (U.S. and Australian) children and their parents. Here, we tested children in the U.S. and in China. We do not specifically predict different findings for these two groups; indeed from an evolutionary perspective, there are reasons to argue that relations between temperament and social cognition should be similar across children from very different cultures. However, conceivably, there may be cross-cultural differences in relations between temperament and ToM, as different cultures value different social behaviors (for review, see Chen & French, 2008). In comparison to Western cultures, China's more collectivist culture values more modest, deferential behavior, and children in that context who are reserved yet socially-attentive may be particularly welcomed to participate in and observe social interactions (Chen & French, 2008; Chen et al., 2011). Chinese culture is particularly averse to reactive or aloof behavior that draws attention to oneself, and children who exhibit such behavior may be excluded from social interaction (Chen & French, 2008; Chen et al., 2011). Because the extent to which one exhibits culturally-valued behavior influences the quality of social interaction in which one participates

and observes, temperament may differentially predict social-cognitive development in the U.S. and China.

In the current study, we focus on specific dimensions of temperament—aggressiveness, social withdrawal, and HPA-axis reactivity—because an emotional reactivity hypothesis provides a focused set of predictions about how these dimensions should relate to ToM. In accord with such a hypothesis, and following prior research (e.g., Wellman et al., 2011), aggression was expected to relate negatively to ToM. Social withdrawal could relate to ToM in two ways. Withdrawal that reflects a calm, interested, yet reserved approach to social interaction should positively relate to ToM, as it did for Wellman and colleagues' (2011). However, if withdrawal is characterized by negative reactivity towards others—reflecting perhaps social fearfulness or annoyance; behavior referred to as *social-avoidance* (Asendorpf, 1990) or *active-isolation* (Harrist et al., 1997)—it should relate negatively to ToM, as it did for Walker (2005). Our use of behavioral and physiological measures of temperamental characteristics should allow us to differentiate between these forms of social withdrawal and, potentially, find contrasting relations with ToM.

HPA-axis response may also relate to ToM, but this has not been directly examined in previous work. High HPA-axis activation has been associated with factors that may undermine ToM development, such as fearfulness, anger, and peer rejection (Donzella et al., 2000; Gunnar et al., 2003; Talge, Donzella, & Gunnar, 2008). Elevated cortisol reactivity has also been found for children who engage in aggressive behavior with peers (Lopez-Duran et al., 2009; Murray-Close et al., 2008), which is predictive of poorer ToM in early childhood (Wellman et al., 2011). However, the role of the HPA axis in aggressive behavior is complex, with some studies demonstrating a negative relation between basal cortisol and antisocial behavior (for meta-

analysis, see Alink et al., 2008). Moderate HPA-axis reactivity has been linked to factors that can facilitate ToM development, including more advanced executive function, increased attention to social stimuli, and greater peer acceptance (Blair, Granger, & Razza, 2005; Gunnar et al., 1997; Lewis & Ramsay, 1997). Thus, it may be that HPA-axis activation predicts more or less advanced ToM development, depending on the level of reactivity and other temperamental characteristics of the child. In the current study, we examine the direct relation between HPA-axis reactivity and ToM and additionally examine how that relation is moderated by children's social withdrawal.

Method

Participants

Preschoolers (ages 44 to 63 months; $M = 53$) from Beijing, China and from a university city in the Midwestern United States participated in a three-day study assessing children's cognitive and behavioral development. Most children came from two-parent, middle-class households, and most parents were college-educated. Most U.S. participants were Caucasian. Of the 117 total children who participated, 102 had data for all three days of the study and all variables focal to the current investigation. This final sample consisted of 51 boys and 51 girls; 55 of whom were from China and 47 from the U.S. These are relatively small samples of convenience that differ in several ways (e.g., number of siblings, variability in executive-function and theory-of-mind performance). Nevertheless, with 102 children overall, the data provide a sensible assessment of certain focal relations between temperament and theory of mind.

Procedures

Parents completed pre-validated measures of children's temperament and behavior, and children were tested on three consecutive days with various laboratory tasks. Each day started with 30 minutes of a calming activity in the laboratory, followed by a 10-minute challenging self-regulation task (see Appendix A), and then 80 minutes of additional calming activities and cognitive tasks. Children completed an executive function (EF) task each day. Additionally, on day 1 children completed false-belief tasks, and on day 3 children's verbal and non-verbal IQ were assessed. We measured physiological reactivity via salivary cortisol and, because cortisol varies along with many contextual factors, collected multiple assays on each day of the study.

Measures

Behavioral measures of temperament. To assess children's temperament and behavior, we had mothers complete two widely-used questionnaires: the Child Behavior Checklist (CBCL/2-3; Achenbach, 1992), and an abbreviated version of the Children's Behavior Questionnaire (CBQ; Rothbart, 1989). For Chinese participants, the CBCL was translated into Mandarin Chinese based on the Hong Kong Cantonese version (Leung et al., 2006); and the CBQ was translated into Mandarin Chinese from the English version (Rothbart, 1989). Both the CBCL and CBQ were translated and back-translated by Mandarin-Cantonese and Mandarin-English bilinguals; discrepancies were discussed and resolved between translators. For the CBCL, mothers rated their child on 99 items that describe behavior over the past two months, using a 3-point scale ("0" = not true; "1" = somewhat or sometimes true; "2" = very true or often true of the child). The CBCL is arranged into several subscales, two of which are focal to this study—aggressive behavior and withdrawn behavior. An *aggression* composite was computed by averaging across 15 items of the aggression subscale (e.g., "Hits others" and "Gets in many

fights”) and 4 other relevant items also from the CBCL—“Wants a lot of attention,” “Destroys things belonging to his/her family or other children” “Hurts animals or people without meaning to,” and “Physically attacks people” ($\alpha = .82$ for China; $\alpha = .83$ for U.S.). A *withdrawal* composite was computed by averaging across the 14 items from the CBCL’s withdrawn subscale (e.g., “Avoids looking others in the eye,” “Doesn’t get involved with others”) ($\alpha = .75$ for China; $\alpha = .70$ for U.S.). These composites are similar to those employed by Wellman et al. (2011).

For the abbreviated CBQ, mothers rated 109 items describing their child’s behavioral reactions to different situations, using a 7-point scale (ranging from “1” = Extremely untrue, to “7” = Extremely true). We used the CBQ to consider temperament dimensions that (given an emotional reactivity hypothesis) should *not* be related to ToM—behavioral inhibition and attention shifting. *Behavioral inhibition* was measured using 12 items from the corresponding CBQ subscale ($\alpha = .64$ for China; $\alpha = .71$ for U.S.), and *attention shifting* was measured using 10 items from the related CBQ subscale ($\alpha = .62$ for China; $\alpha = .75$ for U.S.). Subscale scores were computed by averaging across the items.

HPA-axis reactivity. During each of the three days, children’s salivary cortisol was collected via standard collection procedures (Gunnar & White, 2001). Participants chewed on an unflavored cotton roll for approximately 60 seconds. The roll was inserted into a plastic salivette, which was capped and stored in a household refrigerator for a maximum of 24 hours, when it was centrifuged for 5 minutes and stored at -20 C until assayed. Samples were assayed in duplicate using commercial kits (High Sensitivity Salivary Cortisol Enzyme Immunoassay Kit, Salimetrics). U.S. samples were assayed at a neuro-endocrinology laboratory at the University of Michigan, and Chinese samples were assayed at the biopsychology laboratory at Peking University, using identical Salimetrics kits and assay equipment. To minimize inter-assay

variability, the same (U.S.) technician traveled to China, and thus assayed all samples in both China and the U.S. In addition, 10 duplicate samples were assayed in both locations, with less than 5% variation in observed values between locations.

To examine the magnitude of children's cortisol response over the testing session, cortisol was sampled during the laboratory battery at 9 time points: 0, 10, 20, 30, 40, 50, 60, 75, and 90 minutes. The 9 cortisol samples formed a reactivity curve for each child. To assess children's overall reactivity during the 90-minute session, we used *area under the curve (AUC)*, a commonly-used and validated measure of HPA-axis reactivity (e.g., Pruessner, Kirschbaum, Meinlschmid, & Hellhammer, 2003). For each day, for each interval (e.g., 0-10 minutes) a trapezoid was calculated (the average of the cortisol measure from Time_a and Time_b, multiplied by the duration of the interval). This yielded 8 trapezoids per day, per child. For each day, the 8 trapezoids were added to yield the AUC. Children's AUC for each of the three days were highly inter-correlated (r s range from .66 to .67, p s < .001). Our goal was to obtain a general index of children's physiological reactivity (rather than a measure of children's reactions to particular tasks on particular days), therefore children's physiological reactivity score was calculated as the average AUC across the three days.

Each day, children engaged in a calming activity for a half hour before the testing session began. At the beginning of the testing session each day (Time 0-20), children were introduced to a new person (the interviewer) and were given a challenging/frustrating task (see Appendix A). Thus, children evidenced initially high reactivity (relative to their later reactivity) ($M = .094$ mcg/dL, $SD = .052$), which was likely produced by a combination of mild anticipatory stress prior to starting the tasks and exposure to new people without a parent or other caregiver present (Gunnar, Talge, & Herrera, 2009; Stansbury & Harris, 2000; Zimmermann & Stansbury, 2004),

as well as children's engagement in the challenging/frustrating tasks at the beginning of the session (Vangoozen et al., 1998). This initial reactivity in response to novelty is well-documented (see Gunnar et al., 2009); attempting to induce even more pronounced reactivity might have compromised the ethics of the procedures. Children watched calming videos between Time 20 and 40 each day, during which their reactivity decreased to an average of .083 mcg/dL ($SD = .063$) at Time 40. Children then engaged in other tasks (e.g., ToM, IQ, EF), and the sessions ended before Time 90, at which point most children demonstrated a final decrease in reactivity ($M = .072$ mcg/dL, $SD = .043$). In the context of this across-day, across-time profile, greater (or lesser) AUC reflects greater (or lesser) general physiological reactivity.

Theory of mind. An important milestone in theory of mind (ToM) development is an understanding of false beliefs—an appreciation that others hold beliefs that differ from reality. This milestone is achieved by children worldwide (Liu, Wellman, Tardif, & Sabbagh, 2008; Wellman et al., 2001), and measures of false-belief understanding have been used in many studies exploring individual differences in ToM development among young children (e.g., Ruffman et al., 2002; Slaughter et al., 2002), including Wellman and colleagues' (2011) study on predictive relations between temperament and ToM. We used two standard false-belief tasks: an unexpected-contents task (Perner, Leekam, & Wimmer, 1987) and a switched-location task (Wimmer & Perner, 1983). In both tasks, characters were represented with child-figurines and the appearance of the figurines was matched to the ethnicity of each participant. All characters and props were presented on photographs; thus children saw photographs of the child-figurines, photographs of boxes, and photographs of all other stimuli. Children's *ToM* was calculated as the number of tasks (0-2) for which they correctly attributed a false-belief to story characters.

Control measures. To examine whether relations between temperament and ToM survive after controlling for more general capacities that typically relate to temperament and social-cognitive development (Astington & Jenkins, 1999; Carlson & Moses, 2001; Kagan & Snidman, 2004) and to assess whether temperament indiscriminately predicts multiple types of cognitive development, our analyses included measures of two cognitive competencies that are arguably less closely and strongly related to children's social engagement—executive function and vocabulary. Although vocabulary development, like ToM development, is related to social factors such as parent-child conversation (Feldman et al. 2005), vocabulary also reflects learning from non-social sources; for example, TV and books. Moreover, vocabulary as measured in standard tests is often not about people (words in those tests refer to artifacts, places, and so on), whereas our ToM measure (false-belief understanding) reflects social-cognitive knowledge (about people, actions, and mental states). In the U.S., *vocabulary* was gauged using the Vocabulary subtest of the WPPSI-R (Wechsler, 1989). In China, vocabulary was measured with a Mandarin Chinese version of the Vocabulary subtest of the Stanford-Binet Intelligence Scale (Thorndike, Hagen, & Sattler, 1986), which has been successfully used in previous studies (e.g., McBride-Chang et al., 2008). Vocabulary scores were scaled comparably across countries. *IQ* was calculated as the average of children's standardized vocabulary scores and standardized block-design scores on the WPPSI-R (Wechsler, 1989).

Like ToM and vocabulary, executive function (EF) may contribute to children's social-emotional competence (Riggs, Jahromi, Razza, Dillworth-Bart, & Mueller, 2006). EF also facilitates ToM development (e.g., Carlson & Moses, 2001). EF are more general capacities that are not used to operate on social stimuli alone, and our three EF tasks require that children reason about non-social entities; thus we would expect to not find the same relations between

temperament and EF. To measure executive function, three widely-used tasks were employed—Grass/Snow Stroop, Day/Night Stroop, and Luria’s Hand Game (Carlson & Moses, 2001; Diamond & Taylor, 1996; Luria, Pribram, & Homskaya, 1964). See Appendix B for a description of each task. An executive function (EF) composite was computed by averaging across children’s standardized scores for the three tasks.

Results

Descriptive statistics for the focal variables are presented in Table 1. In these samples, compared with U.S. children, Chinese children exhibited more advanced ToM and EF, were more reactive, withdrawn, and aggressive, and demonstrated lower levels of behavioral inhibition and attention shifting. In the U.S. sample, note that there was more limited variability in two of the focal temperament variables—HPA-axis reactivity and withdrawal. Initial zero-order correlations including data from both the U.S. and China indicated that our measure of HPA-axis reactivity was positively associated with parental ratings of children’s physical aggression ($r = .26, p < .01$), and negatively associated with ratings of behavioral inhibition ($r = -.28, p < .05$); relations consistent with prior findings (e.g., Donzella et al., 2000; Gunnar et al., 1997; Lopez-Duran et al., 2009; Murray-Close et al., 2008). This validates that our measure of physiological reactivity is predictive of children’s everyday reactive behavior. However, HPA-axis reactivity was not significantly correlated with social withdrawal ($r = .15, p = .13$). This is perhaps, as outlined earlier, because parents’ reports of children’s social withdrawal can reflect two very different temperaments: (a) a positive, though initially reserved, approach to social interaction, or (b) negative reactivity towards and avoidance of social interaction. Thus, in our primary analyses predicting ToM, we examine an interaction effect between social withdrawal and physiological reactivity to differentiate these two subtypes of withdrawn behavior.

Temperament Predicts Theory of Mind

Initial regression analyses revealed no significant differences between boys and girls in relations between temperament and ToM, and thus our focal analyses combine data from boys and girls. A hierarchical regression analysis assessed the relations between the focal temperament factors and ToM across the U.S. and Chinese samples. To differentiate children who are reactively-withdrawn versus children who are shy, we included in this analysis an interaction effect between children's social withdrawal and physiological reactivity. As outlined in Table 2, age was entered into the first step of our regression. Probably because of our deliberate restriction of age (to a 19-month range) age was not significantly related to ToM performance. The second step included the temperament variables—withdrawal, aggression, HPA-axis reactivity, and the focal interaction effect, along with two non-social temperament dimensions theoretically unrelated to ToM (behavioral inhibition and attention shifting). These temperament variables significantly predicted ToM— $F\text{-change}(6, 94) = 3.02, p < .01$ —accounting for 16% of the variance in false-belief understanding. Consistent with prior findings, aggression was associated with poorer false-belief understanding. More interestingly, HPA-axis reactivity predicted better false-belief understanding; although this relationship was subsumed under a significant interaction between withdrawal and HPA-axis reactivity.

To explore this interaction effect further, we began with a traditional median split on cortisol response, and found two contrasting, though non-significant, linear trends—withdrawal was related to better ToM among children who were less reactive ($\beta = .27, t = 1.33, ns$), and to poorer ToM among children who were more reactive ($\beta = -.13, t = -.56, ns$). An interaction effect indicates that linear relations significantly differ between some portions of the sample, but not necessarily between the upper and lower halves exactly. Thus, we then re-grouped

participants based on their cortisol response, with group sizes that would leave us with enough statistical power to find significant linear relations for each group (40/60% and 60/40%). In regressions, both of these splits revealed contrasting relations for children high versus low in reactivity, but the linear trends were clearest when comparing participants in the lower 60% (AUC: $M = 4.86$ mcg/dL, $SD = 1.29$; per-moment: $M = .054$ mcg/dL, $SD = .014$) versus participants in the upper 40% (AUC: $M = 10.76$ mcg/dL, $SD = 4.23$; per-moment: $M = .120$ mcg/dL, $SD = .047$) in terms of physiological reactivity. In two separate regressions, withdrawal was positively related to false-belief understanding for less reactive children ($\beta = .44$, $t = 2.16$, $p < .05$); and was negatively related to false-belief understanding for the more-reactive children ($\beta = -.37$, $t = -1.72$, $p = .09$), though this latter trend was only marginally significant. These contrasting linear relations are depicted in Figure 1.

A sizable number of children represented low-withdrawal/low-reactivity ($n=30$); withdrawal/low-reactivity ($n=30$); low-withdrawal/reactivity ($n=17$); and withdrawal/reactivity ($n=25$). Using these groupings allows us to identify other temperamental characteristics that differ between withdrawn children who were reactive and withdrawn children who were less reactive, as assessed by parental reports on the Children's Behavior Questionnaire (Rothbart, 1989). One-way ANOVAs revealed that these four groups differed significantly in perceptual sensitivity ($F(3, 98) = 4.62$, $p < .01$), behavioral inhibition ($F(3, 98) = 5.98$, $p < .001$), and smiling and laughter ($F(3, 98) = 4.96$, $p < .01$). Tukey post-hoc pair-wise comparisons revealed that withdrawn/reactive children had significantly poorer perceptual sensitivity ($M = 4.61$, $SD = 1.08$) than withdrawn/low-reactive children ($M = 5.24$, $SD = .76$), low-withdrawn/low-reactive children ($M = 5.29$, $SD = .87$), and low-withdrawn/reactive children ($M = 5.52$, $SD = .68$). Withdrawn/reactive children also had significantly poorer behavioral inhibition

($M = 4.30$, $SD = .56$) than low-withdrawn/low-reactive children ($M = 4.97$, $SD = .61$), and low-withdrawn/reactive children ($M = 4.62$, $SD = .67$); and withdrawn/reactive children had significantly lower levels of smiling and laughter ($M = 5.28$, $SD = .75$) than low-withdrawn/low-reactive children ($M = 5.93$, $SD = .55$). No significant differences emerged between the withdrawn/low-reactive group and the low-withdrawn groups for these temperament dimensions.

Comparing U.S. and Chinese Samples

The extent to which this general pattern is similar or different across the U.S. and Chinese samples is also of interest, although here conclusions are necessarily more tentative due to smaller samples of children within each country and more limited variability in our focal variables, in particular in the U.S. To assess whether this general pattern of results was carried by either U.S. or Chinese children alone, follow-up regressions were conducted that included participants' country in the first step and interactions between each of the focal temperament variables and location in the second step. In these regressions, none of the interaction terms significantly predicted ToM, suggesting that predictive relations between temperament and ToM did not significantly vary by location. The sole significant relation concerning location was one between location and ToM ($\beta = .24$, $t = 2.47$, $p < .05$). As reported in Table 1, on average, Chinese children outperformed their U.S. counterparts on false-belief understanding.

Lack of significant interaction effects alone is not firm evidence that relations between temperament and ToM are parallel across the two countries, especially given the small sample for each country. To further examine relations for each country, separate regressions were conducted for the U.S. and for China; these results are presented in Table 3. Because the samples for the U.S. and China are relatively small, and each regression contains seven variables, this limited our ability to detect statistically significant relations. Nonetheless, some cross-national

comparisons can be made by examining betas from the two regressions. First, note that the positive relation between HPA-axis reactivity and ToM is evident in both countries, as is the negative relation between aggression and ToM. At the same time, the significant interaction between HPA-axis reactivity and withdrawal is evident only in China. This cross-cultural difference may be accounted for by the fact that there was significantly greater variability in both HPA-axis reactivity and withdrawal in China compared to the U.S. (see Table 1). Thus, the lack of a significant interaction in the U.S. may be an artifact of limited heterogeneity in the data for those key variables. Alternatively, this may reflect genuine differences in the types of social behavior that support social-cognitive development in these two cultural contexts.

Discriminant Analyses

Potentially, many temperament dimensions, not just those encompassed by the emotional reactivity hypothesis, might predict ToM. To assess this, we included in our analyses two other non-social temperament dimensions *not* relevant to the emotional reactivity hypothesis—behavioral inhibition and attention shifting. As illustrated in Table 2, these dimensions did not predict children's ToM. Additional regression analyses assessed whether the focal temperament variables indiscriminately predicted general cognitive development, rather than social-cognitive development in particular. As shown in Table 2, among the temperament variables that predicted ToM, none significantly predicted executive function (EF) or vocabulary. Further, a temperament dimension that did predict EF—behavioral inhibition—did *not* predict ToM.

Finally, to examine whether individual differences in EF and IQ account for the association between temperament and ToM, we conducted a further hierarchical regression analysis predicting ToM, while controlling for age, EF and IQ (an average of children's standardized vocabulary scores and standardized block-design scores on the WPPSI) in the first

step. The temperament variables (aggression, shyness, HPA-axis reactivity, behavioral inhibition, and attention shifting) were entered in the second step. Although IQ ($\beta = .21, t = 1.89, p = .062$) and EF ($\beta = .17, t = 1.67, p = .097$) both related to ToM, the temperament variables collectively continued to predict significant additional variance in ToM— $F\text{-change}(6, 92) = 2.29, p < .05; R^2\text{-change} = .12$. Further, the focal effects of HPA-axis reactivity ($\beta = .22, t = 2.20, p < .05$) and the interaction between HPA-axis reactivity and withdrawal ($\beta = -.24, t = -2.54, p < .05$) remained significant. Aggression was still negatively related to ToM, although non-significantly ($\beta = -.25, t = -1.59, p = .12$).

Discussion

Recent research with non-human animals has inspired a provocative account of how social-cognitive capacities might have evolved in a variety of species—the *emotional reactivity hypothesis*. As first formulated, according to this hypothesis, members of certain animal families—namely canids and the great apes—with less reactive, less aggressive, and more sociable temperaments were better suited to detect, attend to, and respond to others' communicative cues or intentional states (Hare & Tomasello, 2005a; Hare, 2007). Over the course of phylogeny, these temperaments and related social-cognitive abilities were selected either via domestication (for canines) or natural selection (for apes). The current study with human children addresses the emotional reactivity hypothesis from an ontogenetic standpoint. There are several ways in which temperament might influence children's social-cognitive development. Children with less reactive—especially less aggressive—temperaments may be more welcomed in social situations, giving them more opportunities to learn about others. Moreover, those who are less reactive towards others—less socially fearful or anxious—may be better suited to attend to and learn from the nuances of others' social cues. Thus, given humans'

social ecology, a less reactive temperament could facilitate the development of a more sophisticated understanding of others' minds. Using behavioral and physiological measures of temperament, we examined whether individual differences in preschoolers' temperament relate to ToM development. Consistent with prior studies (Capage & Watson, 2001; Wellman et al., 2011) and with the emotional reactivity hypothesis, we found that less aggressive children exhibited more advanced ToM. Further, children who experienced greater physiological reactivity demonstrated better ToM.

At first blush, this finding may appear to run counter to the emotional reactivity hypothesis as first formulated with animals—high levels of reactivity should impede humans' ability to apply their social-cognitive skills (Hare, 2007). However, our samples were non-clinical; children who exhibited 'high' reactivity relative to their peers were still in the normal range (see Blair et al., 2004; Blair et al., 2005). Thus, these children exhibited moderate HPA-axis reactivity, which prior studies demonstrate is related to children's social engagement and social attentiveness (Blair et al., 2004; Blair et al., 2005). As others have suggested, there is likely a curvilinear relationship between cortisol reactivity and adaptive outcomes (Keller, El-Sheikh, Granger, & Buckhalt, 2012). With regard to ToM, moderate cortisol reactivity may increase children's attention to social stimuli (for evidence with infants, see Lewis & Ramsay, 1997); which, in turn, supports social-cognitive development (Dunphy-Lelii, LaBounty, Lane, & Wellman, 2014; Wellman et al., 2008; Wellman et al., 2011).

By including physiological measures of reactivity, we additionally clarify the relation between social withdrawal and social-cognitive development. The key finding was an interaction between physiological reactivity and behavioral ratings of social withdrawal, which predicted ToM development. Among children who were less physiologically reactive, withdrawal

predicted a more advanced ToM. These results are consistent with those of Wellman and colleagues (2011), who found that a shy, more socially-observant temperament predicted U.S. preschoolers' false-belief understanding. These results are also consistent with studies of Chinese preschoolers, demonstrating that more reticent children experience greater peer acceptance (Chen et al., 2011; Chen, DeSouza, Chen, & Wang, 2006), which in turn may foster more quality social interaction and social-cognitive development. However, we also found that, among children who were physiologically reactive, social withdrawal predicted a poorer ToM. These more reactive children's withdrawn behavior may not simply reflect shyness, but rather an intense aversion to social interaction that undermines social-cognitive development. This temperament profile is more consistent with socially-avoidant (Asendorpf, 1990) or actively-isolated behavior (Harrist et al., 1997) that has been identified in previous literature. Relative to the rest of the sample, withdrawn/reactive children exhibited poorer perceptual sensitivity—which may impede their attention to social stimuli and thus hamper social-cognitive development (Wellman et al., 2011)—and exhibited poorer behavioral inhibition and less smiling and laughter, which may reduce how welcomed these children are in social situations. These results accord with the idea that children who exhibit different forms of social withdrawal may evidence quite different social-cognitive skills (Harrist et al., 1997). The interaction between social withdrawal and physiological reactivity also clarifies the significant overall positive relation found between physiological reactivity and ToM—physiological reactivity predicted more advanced ToM among children with low levels of social withdrawal, the majority of our sample.

The combined use of behavioral ratings as well as physiological data allowed us to identify children whose reserved behavior might reflect different temperaments—shy versus actively-isolated/avoidant. However, collecting cortisol data is expensive, and we do not believe

that use of cortisol data is necessary for future studies to address these questions, given validation studies such as the present one which demonstrate clear relationships between physiological and observational measures of temperament. Future studies may address these issues using observational measures that clearly distinguish between shyness and reactive-withdrawal. It is likely difficult for parents and teachers to differentiate between shyness and reactive-withdrawal, and parents are often biased in their interpretations of their children's behavior (Seifer, Sameroff, Dickstein, Schiller, & Hayden, 2004); reactive-withdrawn behavior may be interpreted by parents more positively as shyness. Thus, it would be beneficial for future studies to obtain ratings of shyness and withdrawal from trained raters, rather than parents or teachers.

The focal temperament variables—aggression, shyness/withdrawal, and reactivity—were related to social-cognition and not other cognitive capacities (vocabulary or executive function), providing confirmatory evidence that these particular dimensions of temperament predict social-cognition specifically. The robustness of these relations was further affirmed by the fact that our findings were largely consistent between children from the U.S. and China. In both countries, moderate HPA-axis reactivity was positively related to ToM, whereas aggression was negatively related to ToM. However, the key interaction between HPA-axis reactivity and social withdrawal was only statistically significant in the Chinese sample. Because there was significantly less variability in both withdrawal and HPA-axis reactivity in the U.S. sample, we speculate that the non-significant interaction in the U.S. sample may stem from limited variability in the data. Alternatively, this may reflect true cultural differences in the types of social behavior and temperaments that facilitate social-cognitive development. As Chen and colleagues have explained, in contrast to many Western cultures, in China, children who are reserved yet

socially-attentive may be particularly welcomed to participate in and observe social interactions (Chen et al., 2006; Chen et al., 2011; Chen & French, 2008). In contrast, Chinese children who exhibit behavior that attracts attention (e.g., reactive or defiant behavior) may be excluded from social interaction, and this may hamper their social-cognitive development. Additional studies including children from different cultural contexts, with larger samples from a wider age range, and using a wider array of social-cognitive tasks are necessary before making firm claims about the full generalizability of our findings.

Here we demonstrate that children's temperament is related to individual differences in their false-belief understanding during the preschool years. Such individual differences during the preschool years may have long-term implications, as preschool-age ToM is predictive of children's social competence during the elementary-school years, including the quality of children's discourse with friends and the sophistication of children's conflict resolution strategies (for review, see Astington, 2003). Temperament differences might continue to be influential for social-cognitive development beyond the preschool years. In middle- and late-childhood, children's temperament continues to influence the quality of social interactions in which children engage and through which they learn about others (Chen et al., 2011; Harrist et al., 1997). It would thus be worthwhile to examine whether later-developing social-cognitive competencies are compromised among older children with certain reactive temperaments. Future studies can address this issue by assessing relations between ToM and temperament in older children, using extended ToM scales (e.g., Peterson, Wellman, & Slaughter, 2012; Wellman & Liu, 2004).

It is also important to note that other factors may mediate or moderate associations between temperament and ToM development. For example, parent-child attachment quality and parenting characteristics have been shown to moderate relations between preschoolers'

behavioral inhibition and their HPA-axis reactivity (Kertes, Donzella, Talge, Garvin, Van Ryzin, & Gunnar, 2009; Nachmias, Gunnar, Magelsdoft, Parritz, & Buss, 1996), and parenting characteristics are also predictive of toddlers' HPA-axis reactivity (Hastings, Ruttle, Serbin, Mills, Stack, & Schwartzman, 2011). Moreover, Coplan, Arbeau, and Armer (2008) have demonstrated that characteristics of parents (e.g., their neuroticism) moderate relations between children's shyness and their socio-emotional adjustment. Thus it is important to consider parenting, as well as other social-contextual variables (e.g., life stress; Hunter, Minnis, & Wilson, 2011) in future studies of relations between temperament and ToM.

In sum, by combining behavioral and physiological assessments of children's temperament, we provide cross-cultural evidence supporting an emotional reactivity hypothesis for the ontogeny of human social-cognition. We find that children with certain reactive temperaments—specifically, those who are more aggressive and those who are both socially-withdrawn and physiologically-reactive—evidence poorer social cognition. However, our findings join those of Wellman et al. (2011) in clarifying the boundary conditions for an ontogenetic emotional reactivity hypothesis. First, a certain amount of reactivity (gauged here via salivary cortisol) may actually support social-cognitive development, perhaps by facilitating social engagement and attention to social stimuli. Second, the current findings as well as those from Wellman and colleagues (2011) suggest that certain forms of social disengagement may also foster social-cognitive development. Children who are withdrawn from social interaction yet non-reactive evidence advanced ToM, perhaps because these children still attend to and learn from social interactions, even though they may not actively participate. These findings demonstrate the importance of distinguishing between shy and reactively-withdrawn children in future investigations of relations between temperament and social-cognitive development.

References

- Achenbach T. M. (1992). *Manual for the Child Behavior Checklist/ 2-3 and 1992 Profile*. Burlington, VT: University of Vermont, Department of Psychiatry.
- Alink, L R. A., van, IJzendoorn, M. H., Bakermans-Kranenburg M. J., Mesman, J., Juffer, F., & Koot, H. M. (2008). Cortisol and externalizing behavior in children and adolescents: Mixed meta-analytic evidence for the inverse relation of basal cortisol and cortisol reactivity with externalizing behavior. *Developmental Psychobiology*, *50*, 427–450.
- Asendorpf, J. B. (1990). Beyond social withdrawal: Shyness, unsociability, and peer avoidance. *Human Development*, *33*, 250–259.
- Astington, J. W. (2003). Sometimes necessary, never sufficient: False-belief understanding and social competence. In B. Repacholi & V. Slaughter (Eds.), *Individual differences in theory of mind: Implications for typical and atypical development, Macquarie Monographs in Cognitive Science Series* (pp. 13-38). New York: Psychology Press.
- Astington, J. W., & Jenkins, J. M. (1995). Theory of mind development and social understanding. *Cognition & Emotion*, *9*, 151-165.
- Astington, J. W., & Jenkins, J. M. (1999). A longitudinal study of the relation between language and theory-of-mind development. *Developmental Psychology*, *35*, 1311-1320.
- Astington, J. W. (2005). *Why language matters for theory of mind*. US: Oxford University Press.
- Banerjee R., & Henderson, L. (2001). Social-cognitive factors in childhood social anxiety: A preliminary investigation. *Social Development*, *10*, 558–572.
- Belyaev, D. K. (1979). Destabilizing selection as a factor in domestication. *The Journal of Heredity*, *70*, 301-308.
- Blair, C., Peters, R., & Granger, D. (2004). Physiological and neuropsychological correlates of approach/withdrawal tendencies in preschool: Further examination of the behavioral

- inhibition system/behavioral activation system scales for young children. *Developmental Psychobiology*, *45*, 114-124.
- Blair, C., Granger, D., & Razza, R. P. (2005). Cortisol reactivity is positively related to executive function in preschool children attending Head Start. *Child Development*, *76*, 554-567.
- Capage, L., & Watson, A. C. (2001). Individual differences in theory of mind, aggressive behavior, and social skills in young children. *Early Education & Development*, *12*, 613-628.
- Carlson, S. M., & Moses, L. J. (2001). Individual differences in inhibitory control and children's theory of mind. *Child Development*, *72*, 1032-1053.
- Carpendale, J. I. M., & Lewis. C. (2004). Constructing an understanding of mind: The development of children's social understanding within social interaction. *Behavioral and Brain Sciences*, *27*, 79-96.
- Chen, X., DeSouza, A. T., Chen, H., & Wang, L. (2006). Reticent behavior and experiences in peer interactions in Chinese and Canadian children. *Developmental Psychology*, *42*, 656-665.
- Chen, X., & French, D. C. (2008). Children's social competence in cultural context. *Annual Review of Psychology*, *59*, 591-616.
- Chen, X., Wang, L., & Cao, R. (2011). Sensitivity and unsociability in rural Chinese children: Relations with social, school, and psychological adjustment. *Child Development*, *82*, 1531-1543.
- Clow, A., Hucklebridge, F., Stalder, T., Evans, P., & Thorn, L. (2010). The cortisol awakening response: More than a measure of HPA axis function. *Neuroscience and Biobehavioral Reviews*, *35*, 97-103.

Cole (1986). Children's spontaneous control of facial expression. *Child Development* 57, 1309-1321.

Coplan, R. J., Arbeau, K. A. & Armer, M (2008). Don't fret, be supportive! Maternal characteristics linking child shyness to psychosocial and school adjustment in kindergarten. *Journal of Abnormal Child Psychology*, 36, 359-371.

de Waal, F. (2005). *Our inner ape*. New York, NY: Penguin Group.

Diamond, A. & Taylor, C. (1996). Development of an aspect of executive control: Development of the abilities to remember what I said and to "Do as I say, not as I do." *Developmental Psychobiology*, 29, 315-334.

Donzella, B., Gunnar, M. R., Krueger, W. K., & Alwin, J. (2000). Cortisol and vagal tone responses to competitive challenge in preschoolers: Associations with temperament. *Developmental Psychobiology*, 37, 209-220.

Dunphy-Lelii, S., LaBounty, J., Lane, J. D., & Wellman, H. M. (2014). The social context of infant intention understanding. *Journal of Cognition and Development*, 15, 60-77.

Feldman, H. M., Dale, P. S., Campbell, T. F., Colborn, D. K., Kurs-Lasky, M., Rockette, H. E., & Paradise, J. L. (2005). Concurrent and predictive validity of parent reports of child language at ages 2 and 3 years. *Child Development*, 76, 856-868.

Gunnar, M. R., Sebanc, A. M., Tout, K., Donzella, B., & van Dulmen, M. M. H. (2003). Peer rejection, temperament, and cortisol activity in preschoolers. *Developmental Psychobiology*, 43, 346-368.

Gunnar, M. R., Talge, N. M. & Herrera, A. (2009). Stressor paradigms in developmental studies: What does and does not work to produce mean increases in salivary cortisol. *Psychoneuroendocrinology*, 34, 953-967.

- Gunnar, M. R., Tout, K., de Haan, M., Pierce, S., & Stansbury, K. (1997). Temperament, social competence, and adrenocortical activity in preschoolers. *Developmental Psychobiology, 31*, 65–85.
- Gunnar, M. R., & White, B. P. (2001). Salivary cortisol measures in infant and child assessment. In L.T. Singer & P.S. Zeskind (Eds.), *Biobehavioral assessment of the infant* (pp.167-189). New York, NY: Guilford Press.
- Hare, B. (2007). From nonhuman to human mind: What changed and why? *Current Directions in Psychological Science, 16*, 60-64.
- Hare, B., Brown, M., Williamson, C., & Tomasello, M. (2002). The domestication of social cognition in dogs. *Science, 298*, 1634-1636.
- Hare, B., Melis, A., Woods, V., Hastings, S., & Wrangham, R. (2007). Tolerance allows bonobos to outperform chimpanzees on a cooperative task. *Current Biology, 17*, 619-623.
- Hare, B., & Tomasello, M. (2005a). Human-like social skills in dogs? *Trends in Cognitive Sciences, 9*, 439-444.
- Hare, B., & Tomasello, M. (2005b). The emotional reactivity hypothesis and cognitive evolution: Reply to Miklósi and Topál. *Trends in Cognitive Sciences, 9*, 464-465.
- Harris, P. L. (2006). Social cognition. In W. Damon, R. M. Lerner, D. Kuhn & R. Siegler (Eds.), *Handbook of child psychology: Volume 2: Cognition, perception, and language* (6th ed., pp. 811-858). New York: John Wiley & Sons.
- Harrist, A. W., Zaia, A. F., Bates, J. E., Dodge, K. A., Pettit, G. S. (1997). Subtypes of social withdrawal in early childhood: Sociometric status and social-cognitive differences across four years. *Child Development, 68*, 278-294.

Hastings, P. D., Ruttle, P. L., Serbin, L. A., Mills, R. S., Stack, D. M., & Schwartzman, A. E.

(2011). Adrenocortical responses to strangers in preschoolers: Relations with parenting, temperament, and psychopathology. *Developmental Psychobiology*, *53*, 694-710.

Herrmann, E., Hare, B., Call, J., & Tomasello, M. (2010). Differences in the cognitive skills of bonobos and chimpanzees. *PlosOne*, *5*(8), e12438.

Hunter, A. L., Minnis, H., & Wilson, P. (2011). Altered stress responses in children exposed to early adversity: A systematic review of salivary cortisol studies. *Stress*, *14*, 614-626.

Kagan, J., & Snidman, N. (2004). *The long shadow of temperament*. Cambridge, MA: Harvard University Press.

Kaminski, J., Call, J., & Fischer, J. (2004) Word learning in a domestic dog: evidence for 'fast mapping'. *Science*, *304*, 1682-1683.

Keller, P. S., El-Sheikh, M., Granger, D. A., & Buckhalt, J. A. (2012). Interactions between salivary cortisol and alpha-amylase as predictors of children's cognitive functioning and academic performance. *Physiology & Behavior*, *105*, 987-995.

Kertes, D. A., Donzella, B., Talge, N. M., Garvin, M. C., Van Ryzin, M. J., & Gunnar, M. R. (2009). Inhibited temperament and parent emotional availability differentially predict young children's cortisol responses to novel social and nonsocial events. *Developmental Psychobiology*, *51*, 521-532.

Leung, P. W. L., Kwong, S. L., Tang, C. P., Ho, T. P., Hung, S. F., Lee, C. C., Hong, S. L., Chiu, C. M., & Liu, W. S. (2006). Test-retest reliability and criterion validity of the Chinese version of CBCL, TRF, and YSR. *Journal of Child Psychology and Psychiatry*, *47*, 970-973.

- Lewis, M., & Ramsay, D. S. (1997). Stress reactivity and self-recognition. *Child Development, 68*, 621-629.
- Liu, D., Wellman, H. M., Tardif, T., & Sabbagh, M. A. (2008). Theory of mind development in Chinese children: A meta-analysis of false-belief understanding across cultures and languages. *Developmental Psychology, 44*, 523-531.
- Lopez, N. L., Vazquez, D. M., & Olson, S. L. (2004). An integrative approach to the neurophysiological substrates of social withdrawal and aggression. *Development and Psychopathology, 16*, 69-93.
- Lopez-Duran, N. L., Olson, S. L., Hajal, N. J., Felt, B. T., & Vazquez, D. M. (2009). Hypothalamic pituitary adrenal axis functioning in reactive and proactive aggression in children. *Journal of Abnormal Child Psychology, 37*, 169-182.
- Luria, A. R., Pribram, K. H., & Homskaya, E. D. (1964). An experimental analysis of the behavioral disturbance produced by a left frontal arachnoidal endothelioma (meningioma). *Neuropsychologia, 2*, 257-280.
- McBride-Chang, C., Tardif, T., Cho, J. R., Shu, H. U. A., Fletcher, P., Stokes, S. F., et al. (2008). What's in a word? Morphological awareness and vocabulary knowledge in three languages. *Applied Psycholinguistics, 29*, 437-462.
- Murray-Close, D., Han, G., Cicchetti, D., Crick, N. R., & Rogosch, F. A. (2008). Neuroendocrine regulation and physical and relational aggression: The moderating roles of child maltreatment and gender. *Developmental Psychology, 44*, 1160-1176.
- Nachmias, M., Gunnar, M., Mangelsdorf, S., Parritz, R. H., & Buss, K. (1996). Behavioral inhibition and stress reactivity: The moderating role of attachment security. *Child Development, 67*, 508-522.

- Okamoto-Barth, S., Call, J., & Tomasello, M. (2007). Great apes' understanding of other individuals' line of sight. *Psychological Science, 18*, 462-468.
- Perner, J., Leekam, S. R., & Wimmer, H. (1987). Three-year olds' difficulty understanding false-belief: Representational limitation, lack of knowledge or pragmatic understanding? *British Journal of Developmental Psychology, 5*, 125-137.
- Perner, J., Ruffman, T., & Leekham, S. R. (1994). Theory of mind is contagious: You catch it from your sibs. *Child Development, 65*, 1228-1238.
- Pervanidou, P. & Chrousos, G. P. (2011). Stress and obesity/metabolic syndrome in childhood and adolescence. *International Journal of Pediatric Obesity, 6*, 21-28.
- Peterson, C. C., Wellman, H. M., & Slaughter, V. (2012). The mind behind the message: Advancing theory of mind scales for typically developing children, and those with deafness, autism, or Asperger syndrome. *Child Development, 83*, 469-485.
- Pruessner, J. C., Kirschbaum, C., Meinlschmid, G., & Hellhammer, D. H. (2003). Two formulas for computation of the area under the curve represent measures of total hormone concentration versus time-dependent change. *Psychoneuroendocrinology, 28*, 916-931.
- Riggs, N. R., Jahromi, L. B., Razza, R. P., Dillworth-Bart, J. E., & Mueller, U. (2006). Executive function and the promotion of social-emotional competence. *Journal of Applied Developmental Psychology, 27*, 300-309.
- Rothbart, M. K. (1989). Temperament in childhood: A framework. In G.A. Kohnstamm, J. E. Bates, & M. K. Rothbart (Eds.), *Temperament in childhood* (pp.59-76). New York: Wiley.

- Rothbart, M. K., & Bates, J. E. (2006). Temperament. In N. Eisenberg, W. Damon, & R. M. Lerner (Eds.), *Handbook of child psychology: Vol. 3, Social, emotional, and personality development (6th ed.)* (pp. 99-166). Hoboken, NJ: John Wiley & Sons.
- Rubin, K. H., & Mills, R. S. L. (1988). The many faces of social isolation in childhood. *Journal of Consulting and Clinical Psychology, 56*, 916-924.
- Ruffman, T., Slade, L., & Crowe, E. (2002). The relation between children's and mothers' mental state language and theory of mind understanding. *Child Development, 73*, 734-751.
- Saarni, C. (1984). An observational study of children's attempts to monitor their expressive behavior. *Child Development, 55*, 1504-1513.
- Seifer, R., Sameroff, A., Dickstein, S., Schiller, M., Hayden, L. C. (2004). Your own children are special: Clues to the sources of reporting bias in temperament assessments. *Infant Behavior and Development, 27*, 323-341.
- Slaughter, V., Dennis, M. J., & Pritchard, M. (2002). Theory of mind and peer acceptance in preschool children. *British Journal of Developmental Psychology, 20*, 545-564.
- Soproni, K., Miklósi, A., Topál, J., & Csányi, V. (2001). Comprehension of human communicative signs in pet dogs (*Canis familiaris*). *Journal of Comparative Psychology, 115*, 122-126.
- Stansbury, K. & Harris, M. L. (2000). Individual differences in stress reactions during a peer entry episode: Effects of age, temperament, approach behavior, and self-perceived peer competence. *Journal of Experimental Child Psychology, 76*, 50-63.
- Talge, N. M. Donzella, B., Gunnar, M. R. (2008). Fearful temperament and stress reactivity among preschool-aged children. *Infant and Child Development, 17*, 427-445.

- Tardif, T., Wellman, H. M., & Cheung, K. (2004). False belief understanding in Cantonese-speaking children. *Journal of Child Language, 31*, 779-800.
- Thorndike, R. L., Hagen, E. P., & Sattler, J. M. (1986). *Stanford-Binet Intelligence Scale* (3rd ed.). Chicago: Riverside.
- Trut, L., Oskina, I., & Kharlamova, A. (2009). Animal evolution during domestication: The domesticated fox as a model. *Bioessays, 31*, 349-360.
- van Goozen, S. H. M., Matthys, W., Cohen-Kettenis, P. T., Wied, C. G., Wiegant, V. M. & van Engeland, H. (1998). Salivary cortisol and cardiovascular activity during stress in oppositional-defiant disorder boys and normal control. *Biological Psychiatry, 43*, 531-539.
- Voigt, J. (2005). Artificial Intelligence Lab, College of Engineering, University of Michigan, Ann Arbor, MI.
- Walker, S. (2005). Gender differences in the relationship between young children's peer-related social competence and individual differences in theory of mind. *Journal of Genetic Psychology, 166*, 297-312.
- Wang, L., Tardif, T., & Olson, S. L. (2005). Socialization of emotion in preschoolers. *Unpublished manuscript*. University of Michigan.
- Wechsler, D. (1989). *Preschool and Primary Scale of Intelligence-Revised*. San Antonio, TX: Psychological Corporation.
- Wellman, H. M. (1990). *The child's theory of mind*. Cambridge, MA: MIT.
- Wellman, H. M., Fang, F., Liu, D., Zhu, L., & Liu, G. (2006). Scaling of theory-of-Mind understandings in Chinese children. *Psychological Science, 17*, 1075-1081.

- Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development: The truth about false belief. *Child Development, 72*, 655-684.
- Wellman, H. M., Lane, J. D., LaBounty, J., & Olson, S. L. (2011). Observant, nonaggressive temperament predicts theory of mind development. *Developmental Science, 14*, 319-326.
- Wellman, H. M., & Liu, D. (2004). Scaling of theory-of-mind tasks. *Child Development, 75*, 523-541.
- Wellman, H. M., Lopez-Duran, S., LaBounty, J., & Hamilton, B. (2008). Infant attention to intentional action predicts preschool theory of mind. *Developmental Psychology, 44*, 618-623.
- Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition, 13*, 103-128.
- Zimmermann, L. K. & Stansbury, K. (2004). The influence of emotion regulation, level of shyness, and habituation on the neuroendocrine response of three-year-old children. *Psychoneuroendocrinology, 29*, 973-982.

Table 1

Descriptive Statistics for Focal Variables for Children in the U.S. and China

Variable	U.S.	China	T-value
	<i>M (SD)</i>	<i>M (SD)</i>	
Age	53.62 (4.83)	52.47 (3.31)	^a 1.37
Theory-of-Mind	.45 (.65)	.76 (.72)	-2.31 *
HPA-axis reactivity	5.59 (2.72)	8.75 (4.51)	^a -4.35 ***
Withdrawal	.19 (.16)	.38 (.25)	^a -4.54 ***
Aggression	.33 (.25)	.51 (.28)	-3.39 ***
Behavioral Inhibition	4.90 (.75)	4.62 (.68)	2.04 *
Attention Shifting	4.13 (.89)	3.77 (.68)	2.30 *
Executive Function	-.12 (.58)	.09 (.40)	^a -2.03 *

Note. ^aThese *t*-tests are corrected for unequal variances between groups; Levene's Tests for Equality of Variances $F_s > 8.00$, $p_s < .01$.

Age is reported in months.

HPA-axis reactivity is the total area under children's cortisol reactivity curve across the 90-minute testing session, reported in mcg/dl.

Executive Function scores were standardized across both countries, with a mean of 0. Vocabulary was standardized within each country, and thus mean differences in vocabulary between U.S. and Chinese participants cannot be examined.

Table 2

Temperament Predicting Theory of Mind, Executive Function, and Vocabulary, Across U.S. and Chinese Samples

Criterion	ToM		EF		Vocabulary	
	B (SE B)	β	B (SE B)	β	B (SE B)	β
Step 1						
Age	.02 (.02)	.09	.00 (.01)	.03	.06 (.02)	.25 *
Step 2						
HPA-axis reactivity	.05 (.02)	.27 **	.01 (.01)	.11	.03 (.02)	.11
Withdrawal	.24 (.43)	.11	.56 (.31)	.26	-.23 (.59)	-.06
Aggression	-.86 (.40)	-.34 *	-.47 (.29)	-.26	-.89 (.56)	-.26
Behavioral inhibition	-.10 (.12)	-.10	.24 (.09)	.34 **	-.26 (.17)	-.19
Attention shifting	-.09 (.10)	-.10	-.13 (.07)	-.21	.09 (.14)	.08
HPA-axis reactivity X withdrawal	-.23 (.09)	-.26 **	-.03 (.06)	-.05	-.07 (.12)	-.06

Note. ToM = Theory of mind

EF = Executive function

* $p \leq .05$, ** $p \leq .01$

Table 3

Temperament Predicting Theory of Mind in the U.S. and in China

	U.S.			China		
	B (SE B)	β	<i>p</i> -value	B (SE B)	β	<i>p</i> -value
Step 1						
Age	.01 (.02)	.11	.46	.03 (.03)	.15	.28
Step 2						
HPA-axis reactivity	.06 (.04)	.23	.19	.03 (.02)	.20	.13
Withdrawal	.21 (.99)	.05	.83	.40 (.60)	.14	.51
Aggression	-.76 (.59)	-.29	.20	-1.21 (.61)	-.48	.06
Behavioral inhibition	-.03 (.18)	-.04	.86	-.21 (.18)	-.19	.26
Attention shifting	-.01 (.14)	-.02	.93	-.18 (.16)	-.17	.28
HPA-axis reactivity X withdrawal	-.08 (.24)	-.06	.76	-.25 (.10)	-.33	.01

Note. Because of small sample sizes within each country, *p*-values are listed for all variables entered into the regressions.

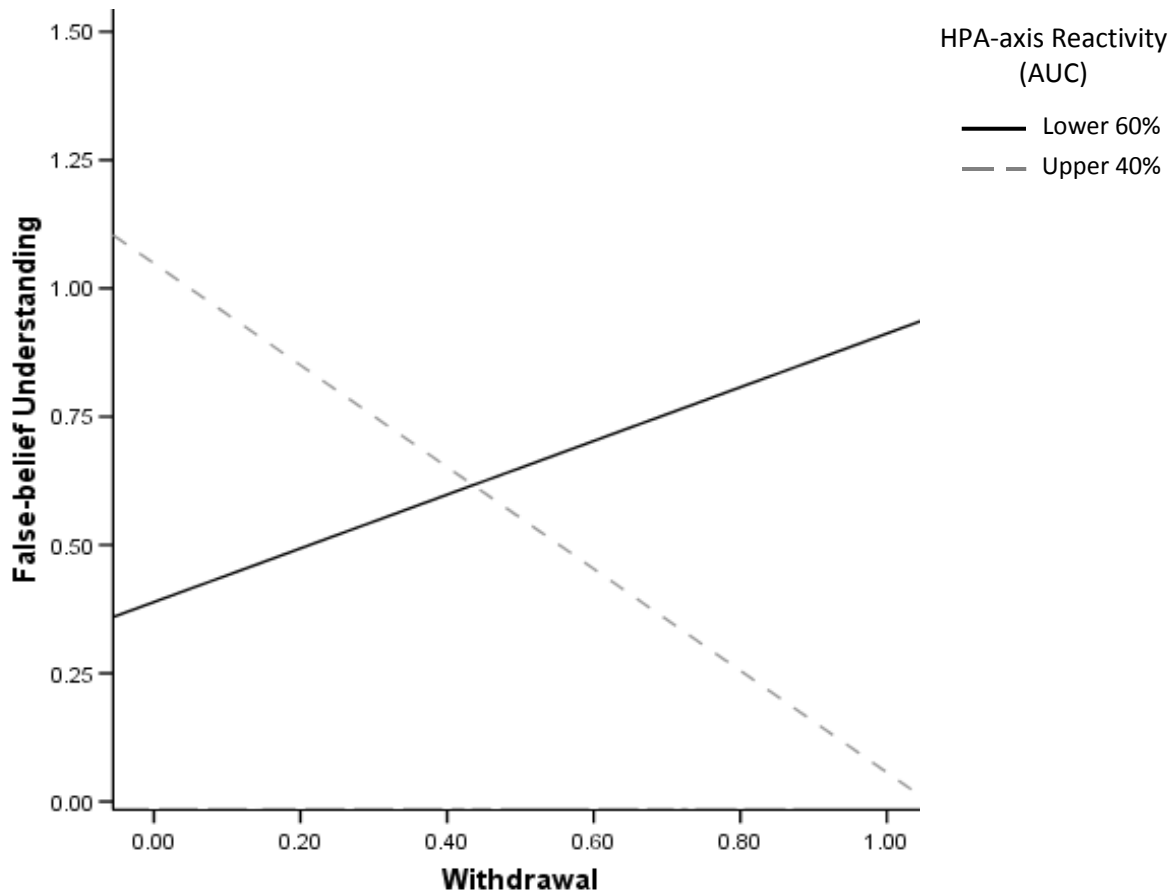


Figure 1. Interaction between physiological reactivity (HPA-axis reactivity) and social withdrawal predicting false-belief understanding. Depicted are fitted regression lines for children who scored within the lowest 60% for reactivity (solid black line) and children who scored within the highest 40% for reactivity (hashed grey line).

Appendix A

Self Regulation Challenge Tasks

Prize Task (day 1). Following the procedures of Saarni (1984) and Cole (1986), children first rank-order a series of potential “prizes” together with an adult (basal phase). Then, the experimenter leaves with the prizes. A new adult enters the room and gives children the least-preferred prize (regulation phase). The new adult remains present and unresponsive for 60 seconds and then exits, leaving children alone in the room for an additional minute. Finally, the first adult returns to the room, asks children how they felt about receiving the prize, and then tells them a mistake was made and offers the most highly-ranked prize (resolution phase). The entire task lasts 5 to 8 minutes, depending on children’s compliance in the initial (basal) phase.

Computer Task (day 2 or 3). Children are shown a computer game (Voigt, 2005) in which they are asked to “lasso” cattle that stray from a path leading to corrals in a barn. The game is relatively easy and children are allowed to try it. Children are told that if they “win” the game, they can pick a prize from a grab-bag, but if they “lose” 3 cattle into the stream, they lose the game and receive no prize. Children then play the game for a few minutes while the experimenter is away from the room. The game works fine until 4 out of 5 cattle are in their corrals, when the button used to “lasso” the cattle stops working. After the original experimenter returns, a second experimenter enters and says that the game is broken. The first experimenter apologizes and children are then given a chance to pick a prize from the grab-bag.

Envelope Task (day 2 or 3). Children are shown a pile of envelopes and a pile of papers, and are instructed to stuff each envelope with one sheet of paper while the experimenter leaves the room. Children are told that if they help the experimenter by stuffing as many envelopes as they can, they may choose a prize from an attractive grab-bag (Wang, Tardif, & Olson, 2005).

Appendix B

Executive Function Tasks

Grass/Snow Stroop. For the Grass/Snow task (Carlson & Moses, 2001), the experimenter shows the child a board with a white square and a green square, and makes sure the child knows the color of grass and snow. During the imitation phase, the child is instructed to point to the white square when the experimenter says “snow” and the green square when the experimenter says “grass”, for 10 trials. During the conflict phase, the child is instructed to point to the square with the opposite kind of color for 10 trials. Children earned 1 point for each correct trial (possible range: 0-20).

Day/Night Stroop. For the Day/Night Stroop task (Diamond & Taylor, 1996), the experimenter has a book of pictures that depict daytime scenes or nighttime scenes. During the imitative phase, the child is supposed to say “Day” when the experimenter reveals a daytime scene, and “Night” when the experimenter reveals a nighttime for 10 trials. During the conflict phase, the child is instructed to say “Night” when the experimenter reveals a daytime scene, and “Day” when the experimenter reveals a nighttime scene for 10 trials. Children earned 1 point for each correct trial (possible range: 0-20).

Luria’s Hand Game. For Luria’s Hand Game (Luria et al., 1964), the experimenter points with either one or two fingers. During the imitative phase, the child is instructed to point in the same manner as the experimenter (i.e., using one or two fingers) for 10 trials. During the conflict phase, the child is instructed to point in the opposite manner from the experimenter (i.e., using one finger if the experimenter used two fingers, and vice versa) for 10 trials. Children earned 1 point for each correct trial (possible range: 0-20).