GUILT, EMPATHY, AND COMPLIANCE IN A NATURALISTIC MORAL SCENARIO: PREDICTING PROSOCIAL AND EXTERNALIZING BEHAVIOR IN 3-7-YEAR-OLD CHILDREN

By

Caitlin J. Listro

A DISSERTATION

Submitted to Michigan State University in partial fulfillment of the requirements for the degree of

Psychology—Doctor of Philosophy

ABSTRACT

GUILT, EMPATHY, AND COMPLIANCE IN A NATURALISTIC MORAL SCENARIO: PREDICTING PROSOCIAL AND EXTERNALIZING BEHAVIOR IN 3-7-YEAR-OLD CHILDREN

By

Caitlin J. Listro

Current science offers only limited answers concerning the development of empathy disturbances. Indeed, few studies have attempted to empirically identify the developmental trajectory of empathy to define either normal or aberrant developmental patterns. The present study aimed to use an observational approach to assess empathy, guilt, and obedience in children, and to validate this approach by exploring how these observational measures of child moral behavior associated with characteristics theoretically linked to moral development and antisocial behavior. We utilized a videotaped Picture Tearing task in which the child is presented with a moral dilemma. Trained coders rated the tasks for several child behaviors (e.g. guilt, gaze avoidance, defiance) using a coding scheme adapted from the Lab-TAB (Goldsmith et al., 1993). Variations in moral behavior were investigated using person-centered (cluster analysis) and variable-centered (factor analysis) methods, then associations between resulting behaviors and other relevant child characteristics (temperament, externalizing behaviors) were examined concurrently and over time. In general, results indicated that empathic verbalizations and defiance were consistently associated with externalizing pathology. This association was observed concurrently; empathy did not predict externalizing over time. Overall, these results suggest that compliant without complaint is the most adaptive response at this age. Furthermore, the Picture Tearing task does provide useful data about empathic behavior and its associations in

young children. Recommendations are made for adaptations to the task and coding scheme to improve the measurement of moral behavior in future research.

Keywords: guilt, empathy, compliance, moral emotions, moral behavior

ACKNOWLEDGEMENTS

It is difficult to quantify the importance of other people in the process of proposing, creating, and passing a dissertation project. They are invaluable, essential. First and foremost, I would like to thank my adviser, Emily Durbin, without whom I might be somewhere else now, in another life with only a master's degree and an uncertain career path. I cannot adequately express what her support has meant to me. I am also indebted to my committee members, Alytia Levendosky, Brooke Ingersoll, and Jenna Neal, for their time and thoughtful consideration. I am doubly indebted to David Clark and Allison Gornik, my wonderful labmates, who provided emotional and statistical assistance; to Taryn Smith, who was integral in the development of the coding scheme; and to Evan Tacey, Madeleine Lenhausen, and Amanda Kilgore, who all logged countless hours coding videos of children being asked to do patently horrible things.

A world of thanks to my amazing family, my parents Christine and Sammy Listro and Sarah, Debbie, and Jim Posey. You are my world. To my friends and cohort mates who sat up long hours with me or just made me smile through the hardest parts: Katey Smagur, Britny Hildebrant, Aurora Dixon, Amber Mandalari, David Johnson, Anna and Andrew Lennard, Allison Gornik, Nicola Bernard, Erika Vitale, Alana Harrison, Rachael Goodman.

To my dog, Bailey, whose positivity always energizes me. And to my cat, Nemo, who only occasionally complained while I was in the midst of dissertation chaos.

iv

TABLE OF CONTENTS

LIST OF TABLES i	X
LIST OF FIGURES	x
INTRODUCTION	1
The Development of Empathy	3
Historical Models of Empathy Development	3
Prototypical Timeline for the Development of Empathic Skills in Early Childhood	d
	4
Empathic Skills and The Mechanisms Underlying Their Development	8
Affective Empathy: The Role of Arousal	9
Dispositional differences in arousal.	9
Environmental modulation of arousal.	1
Cognitive Empathy: Perspective Taking 1	2
Emotion recognition. 1	2
The Social Environment: Obedience and Compliance 1	4
A History of the Scientific Study of Obedience 1	6
Obedience in Children 1	9
Changes in Compliance Across Development 2	,4
The Present Study 2	.5
Aims of the Present Study 2	; [
Overarching Aim: Validate the picture tearing coding scheme as an assessment of)†
child moral development characteristics.	, /
Specific Subordinate Aims 2	,δ
Aim 1: Describe variations in empatitic responding in a morally relevant	0
Scenario.	0
AIII 2: Examine associations between observed moral behavior and child	0
Aim 2: Examina predictive associations between compliance, arousel, and	9
empathy and later externalizing behaviors.	0
METHOD	1
METHOD 3	1
Overview of Design 3	1
Participants 3	2
Exploring New Objects (Durbin, 2010: Easr, happings)	2
Making a t shirt (Durbin, 2010; engagement, happiness).	3
Stranger approach (Lab_TAB: Goldsmith et al. 1993: Fear)	3
Impossibly perfect green circles (Lab-TAB: Goldsmith et al. 1993; Anger	5
sadness)	3
Popping hubbles (Lab-TAB: Goldsmith et al. 1993: Activity happiness) 3	3
Diorama snakes (Fear)	3
Snack delay (Lab-TAB; Goldsmith et al., 1993; Effortful control).	4

Picture tearing (Goldsmith et al., 1995; Guilt, empathy, sadness, compliance).	34
Balloon bop (Goldsmith & Rothbart, 1996; Effortful control, happiness).	35
Transparent box (Lab-TAB: Goldsmith et al., 1993; Anger, sadness).	35
Simon says (Strommen, 1973; Effortful control).	35
Tell a story (Durbin et al., 2007: Fear).	35
Pop-up snakes (Lab-TAB: Goldsmith et al., 1993: Anticipatory positive affect.	00
happiness surprise).	36
Walk-a-line slowly (Kochanska et al., 1996; Effortful control).	36
Box empty (Lab-TAB: Goldsmith et al., 1993: Anticipatory positive affect, and	per.
sadness).	36
Coding of Child Temperament Laboratory Tasks	36
Observational coding of picture tearing task.	36
Observational coding of laboratory tasks.	37
Other Tasks	39
Emotion recognition training.	39
Peabody Picture Vocabulary Test (PPVT).	40
Measures	40
Child Behavior Checklist (CBCL).	40
Children's Behavior Questionnaire (CBQ).	42
Experimenter ratings of child traits.	42
Alabama Parenting Ouestionnaire (APO).	43
Data Analytic Plan	43
Aim 1: Describe variations empathic responding in a morally relevant scenario	Э.
	43
Factor analysis.	43
Cluster analysis.	44
Aim 2: Examine associations between observed moral behavior and child	
characteristics theoretically linked to moral development and empathy.	45
Demographic Variables.	48
Cognitive characteristics.	50
Temperamental characteristics.	50
Parenting factors.	53
Aim 3: Examine predictive associations between compliance, arousal, and	
empathy and later externalizing behaviors.	54
RESULTS	56
Recoding of Picture Tearing Variables	56
Bivariate Correlations Among Recoded Variables	57
Associations Between Task Behaviors and Demographic Variables and Environme	ent
Characteristics	59
Age and sex.	60
Cognitive characteristics.	61
Parenting characteristics.	61
Overall.	62
Exploratory Factor Analysis	63

Confirmatory Factor Analysis of Empirically Derived Factors	64
Confirmatory Factor Analysis of Rationally Derived Factors	65
Final Factor Analysis	66
Summary of Factor Analysis	67
Associations Between Picture Tearing Task Composites and Other Child and	
Environment Characteristics	67
Demographic variables.	68
Cognitive characteristics.	68
Parenting variables.	68
Overall.	69
Derivation of Clusters	69
Hierarchical cluster analysis.	69
K-means cluster analysis.	71
Description of clusters in three cluster solution.	71
Cluster 1: Moderate compliers.	72
Cluster 2: Low compliers.	72
Cluster 3: High compliers.	72
External validation.	73
Demographics.	74
IQ and emotional intelligence.	74
Temperament traits.	75
Problem behaviors.	76
Parenting behaviors.	76
Overall.	76
Gender of Victim as a Predictor of Child Task Behaviors	77
Prediction of Concurrent Problem Behaviors Using Composites Derived from Factor	
Analysis	78
Correlations between independent variables and dependent outcomes.	78
Picture tearing composites.	78
Demographic variables.	78
Verbal and emotional ability.	79
Temperament traits.	79
Parenting characteristics.	80
Overall.	80
Hierarchical regressions predicting problem behaviors from task behavior composites.	
	81
Hierarchical regressions controlling for other child and environment characteristics.	
	81
Summary of hierarchical regressions using task composites.	82
Prediction of Concurrent Problem Behaviors Using Individual Task Variables	83
Correlations between independent variables and dependent outcomes.	83
Hierarchical regressions predicting problem behaviors from task behavior.	83
Hierarchical regressions controlling for other child and environment	<u> </u>
characteristics.	84
Summary of hierarchical regressions using individual task variables.	84
Prediction of Concurrent Temperament Traits Using Picture Tearing Variables	85

Correlations between independent variables and dependent outcomes.	85
Hierarchical regressions predicting temperament traits from task behavior.	87
Summary of regressions predicting temperament traits from task behaviors.	90
Longitudinal Prediction of Problem Behaviors Using Composites Derived from Factor	
Analysis	92
Summary of longitudinal prediction of problem behaviors.	94

DISCUSSION	95
Confirmation and Extension of Previous and Hypothesized Findings	96
Unexpected Findings	106
Conclusions and Areas for Future Study	112
Validity of the picture tearing coding system and recommendations for fut	ure use.
	112
Limitations and future applications.	116
Implications for the study of moral development.	117

APPENDICES	126
APPENDIX A: FIGURES	127
APPENDIX B: TABLES	137
APPENDIX C: Original picture tearing coding scheme	181
APPENDIX D: Description of clusters in four cluster solution	185
APPENDIX E: Recommended final picture tearing coding scheme	187
REFERENCES	192

REFERENCES

LIST OF TABLES

Table 1. Intraclass correlation coefficients for video coding variables	139
Table 2. Picture Tearing codes relevant to each epoch of the task	140
Table 3. Comprehensive list of hypotheses	142
Table 4. Hypothesized results of factor analysis	148
Table 5. Hypothesized associations for variables investigated at baseline	149
Table 6. Descriptive statistics for picture tearing variables	150
Table 7. Descriptive statistics for recoded picture tearing variables	152
Table 8. Bivariate correlations among picture tearing variables	154
Table 9. Correlations between picture tearing variables and child characteristics	156
Table 10. Correlations between picture tearing variables and parenting characteristics	158
Table 11. Factor structures based on exploratory factor analysis	161
Table 12. Fit statistics and comparative analyses for confirmatory factor analyses	162
Table 13. Factor loading estimates for final five factor model	163
Table 14. Descriptive statistics for three cluster solution	164
Table 15. Correlations between task variables, child and environmental characteristics, a problem behaviors	und 166
Table 16. Correlations between picture tearing variables and temperament traits	170
Table 17. Hierarchical multiple regression analyses predicting current child problem ber from picture tearing task composite variables	naviors 173
Table 18. Hierarchical multiple regression analyses predicting current child problem beh from single picture tearing task variables	aviors 176
Table 19. Hierarchical multiple regression analyses predicting current child temperament from picture tearing task composite variables	t traits 177

LIST OF FIGURES

Figure 1. Diagram of moral development and its relationship to child internal and externa characteristics and secondary outcomes	ıl 128
Figure 2. Chart of Video Progression and Relevant Codes	129
Figure 3. Hierarchical cluster analysis agglomeration coefficients plotted against stage of analysis	130
Figure 4. Gap statistic for hierarchical cluster analysis graphed against number of cluster	s 131
Figure 5. Differences between cluster means for three cluster solutions	132

INTRODUCTION

In the wake of the Holocaust, Stanley Milgram set out to test the prevailing public notion that the Nazi officers who perpetrated genocide were simply evil. He famously found that, when commanded by an experimenter, the vast majority of adults in his study would obey orders to electrically shock another person using what they believed were dangerous levels of voltage (Milgram, 1965). As Milgram demonstrated, actual moral or immoral behavior is not solely determined by one's internalized morals. In understanding antisocial behavior, it is thus crucial to understand the processes by which morality develops and factors that predict moral decisions in various situations. It is also crucial to understand the qualities that distinguish typical moral development from aberrant development, and subsequently the processes by which aberrant development of factors related to moral behavior occurs. In the child literature, this problem is especially relevant to understanding children who appear to lack moral feelings altogether: those with Callous-Unemotional (CU) Traits.

CU traits in children encompass a constellation of features characterized by lack of guilt, lack of empathy, callous use of others for one's own gain, and underreactivity to threatening and emotionally distressing stimuli (Frick et al., 2003; Frick & Morris, 2004; Frick, Ray, Thornton, & Kahn, 2014; Frick & White, 2008). Children with CU traits are unique among antisocial youth in that they display blunted affectivity, absent remorse, and severe impairments in empathy. These empathy impairments are severe compared not only to healthy controls, but also to other individuals with antisocial behavior problems (e.g. de Wied, van Boxtel, Matthys, and Meeus, 2011). Even in children who may not reach threshold for CU traits, low empathy is robustly and strongly associated with antisocial behavior and violence (e.g. Asendorpf &

Nunner-Winkler, 1992; Hastings et al., 2000; Jolliffe & Farrington, 2004; Krettenauer, Asendorpf, & Nunner-Winkler, 2013; Lovett & Sheffield, 2007).

Despite the preponderance of research linking impaired empathy to aggression and violence, current science offers only limited answers concerning the development of empathy disturbances. Indeed, few studies have attempted to empirically identify the developmental trajectory of empathy to define either normal or aberrant developmental patterns. Moral development and decision making is also extremely complex (see Figure 1). It depends not only on moral emotions such as empathy, but also on cognitive skills and external influences, such as the actions of victims or authority figures. The current literature provides a limited understanding of how these factors interrelate to produce behaviors in observable moral situations. The present study aims to use an observational approach to assess empathy, guilt, and obedience in children, and to validate this approach by exploring how these observational measures of child moral behavior associated with characteristics theoretically linked to moral development and antisocial behavior, including child temperament traits and externalizing problems (both concurrently and prospectively) and aspects of the parent-child relationship. We utilized a videotaped task in which the child is presented with a moral dilemma: obey the adult experimenter and destroy someone's "cherished" item, or preserve the item and defy the adult authority. In this way, we hoped to clarify individual differences in empathy and obedience as markers of moral behavior, and to test for the construct validity of these observed individual differences by quantifying their association with factors presumed to play a causal role in children's moral development and with outcomes presumed to mark abnormal (versus normal) moral development, including antisocial behavior and callous-unemotional traits.

The Development of Empathy

Historical Models of Empathy Development

Initial approaches to moral development delineated universal processes that were thought to unfold along approximately the same timeline and the same stages for every child. Both Piaget (1965) and Kohlberg (Kohlberg & Kramer, 1969) suggested that children first demonstrate simplistic rule-responsiveness based on the dictates of authority figures and the threat of punishment, and only later learn to think independently and relativistically about issues of morality and to develop personal values to draw upon when engaging in moral decisionmaking. These models viewed young children as pragmatic, egocentric, pre-moral agents responding to external strictures with relative insensibility to the reasoning behind their actions (Zahn-Waxler, Radke-Yarrow, Wagner, & Chapman, 1992). They also identified the emergence of empathy or guilt (critical components of moral awareness and moral behavior) as not occurring until late childhood at the earliest. However, empirical research suggests that primitive forms of moral emotions (i.e. empathy, guilt) can be observed in children as young as 12 months of age (Zahn-Waxler, Radke-Yarrow, Wagner, & Chapman, 1992). Indeed, children as young as 3 months of age are capable of recognizing goal-directed activity, even in inanimate objects (Hamlin, Wynn, & Bloom, 2010; Shimizu & Johnson, 2004), and demonstrate aversion to objects perceived as acting antisocially, but preference towards objects perceived as acting prosocially (Hamlin, Wynn, & Bloom, 2010). Thus, while they may not be able to articulate moral principles in a sophisticated manner, they still demonstrate a basic awareness of 'right' and 'wrong' and the ability to evaluate personified entities based on these principles.

In contrast with early stage models of Piaget and Kohlberg, modern theories of moral development describe a fluid trajectory that begins at birth (Hoffman, 1982; Kochanska, 1993). Both theory and empirical work suggest that this process does not occur on a fixed timeline with

invariant stages, but rather, is affected by continual transactions between the child's disposition and his or her environment, producing individual differences in important markers of moral behavior and reasoning. Modern empathy theory asserts that, at the most basic level, humans are biologically prepared to develop empathy (Hoffman, 1982). Hoffman (1982) suggested that children develop empathy first in infancy as mimicry of others' affective expressions, and subsequently develop the internalization of rule-oriented messages delivered by caregivers or parents. "Internalization" is the key process in this scenario, by which the child accepts parental strictures as their own beliefs. Eventually, the child bases their conduct on intrinsic, absorbed messages rather than external threats of punishment (Rose, 1999). While the progression of development described here is similar to that proposed by Piaget (1965) and Kohlberg (Kohlberg & Kramer, 1969), it differs in that it does not view children as purely authority-driven and insensitive to higher levels of moral reasoning. Rather, it posits that children internalize the moral messages suggested by adult rules (e.g. do not hit your sister because it hurts her) rather than simply following rules without awareness of the reasoning behind them (Zahn-Waxler, Radke-Yarrow, Wagner, & Chapman, 1992).

Prototypical Timeline for the Development of Empathic Skills in Early Childhood

The hypothesis that empathy in humans is biologically predisposed is supported by a wealth of evidence concerning the prototypical development of empathy. While research on the empathic development in preverbal children is limited by methodological constraints, available studies support that the development of empathic skills occurs as soon as a child is born. Even infants 2-3 days old exhibit reflexive crying upon hearing the cries of other infants (Field, Woodson, Greenberg, & Cohen, 1982; Sagi & Hoffman, 1976). Reflexive crying may reflect a biological disposition for infants to respond to the distress of others. It is also unlikely that

infants are merely responding to a noxious stimulus (the sound of crying) with distress, given that the infants preferentially cried in response to other infants' crying in comparison not only to their own pre-recorded cries, but also to other loud sounds (Field, Woodson, Greenberg, & Cohen, 1982). These data support the contention that infants are predisposed to attend to and to exhibit 'empathic' responses to the distress of others (Sagi & Hoffman, 1976).

In further support of this notion, studies of infants have found that infants displayed negative emotional arousal in response to the negative affective states of their depressed mothers (Cohn, Campbell, Matias, & Hopkins, 1989), to their mothers' simulated distress, and to videos of distressed peers (Roth-Hanania, Davidov, & Zahn-Waxler, 2011) From a developmental standpoint, these data indicate that infants are aware of and responsive to the distress of others much earlier than previously believed.

Expressions of empathy analogous to those found in adults are first identifiable around 18 months of age, with some expressions of mature empathy observable around 24 months (Belacchi & Farina, 2012). During this period, children learn the function of symbols and are increasingly capable of representational thought; they begin to understand hypotheticals and to mentally represent complex information (e.g. Bruner, 1972; Piaget, 1932). The onset of these skills facilitates nascent perspective taking abilities, which are predicated on the ability to mentally represent the emotions and desires of others. Evidence for perspective taking has been observed in the form of toddlers' responses to others' distress (Zahn-Waxler, et al., 1992). Zahn-Waxler and colleagues (1992) demonstrated that many children in their sample as young as twelve months exhibited attempts to comfort another person in distress; by 24 months, almost all children exhibited comforting attempts. These responses, while relatively simplistic, suggest a

form of empathic role taking, in that children are recognizing the other person's distress and are responding as they might like to be comforted in a similar situation (Zahn-Waxler et al., 1992).

The age of approximately 24 months is a particularly important period in empathy development. At this time, the child is developing both increased autonomy and sophisticated differentiation of self and other (Zahn-Waxler et al., 1992), which may enable them to distinguish between their own desires and goals and those of another person when these are in conflict. Children at this age are also developing the emotional capacity to recognize and experience others' emotional states, the cognitive capacity to interpret others' emotional states, and the behavioral capacity to identify appropriate responses and to act (Zahn-Waxler et al., 1992). Consistent with this notion, in the study of toddlers discussed above, increasing age was related to a greater sophistication, variety, and appropriateness of comforting attempts (Zahn-Waxler et al., 1992). Simply put, as children developed a mature theory of mind, they were able to identify comforting behaviors tailored to the other person, presumably via rudimentary perspective taking.

Increased autonomy and mobility also facilitate the processes of socialization and the internalization of parental restrictions (Kochanska, 1993). For the first time, children are able to perform behaviors that may run counter to parental desires, some of which will incur parental prohibitions. Consequently, expressions of guilt following a transgression also emerge around this time (e.g. Aksan & Kochanska, 2005; Kochanska & Aksan, 2006). While difficult to assess, guilt in children has typically been operationalized via behavioral indicators, including: gaze avoidance, apologizing, bodily tension, negative affect, and self-blame (Kochanska & Aksan, 2006). Guilt generally results from empathic awareness of another's distress in concert with the awareness that one has caused this distress via a transgression (Hoffman, 1975). Thus, guilt is an

important epiphenomenon of empathy. While measurement limitations prohibit clarity in the question of whether children at this time feel empathy per se, the aforementioned research suggests that at least primitive empathy characterized by an affective component (i.e. recognizing the other's distress, manifesting guilt) and a cognitive component (i.e. identifying what the other might desire as comfort in the situation) can be observed in children of this age.

The next critical milestone occurs approximately between 24 and 36 months, when children first demonstrate an ability to obey rules in the absence of a parent (Kochanska, 1993). Prior to this milestone, children typically desist in transgressing only because a caregiver is present to induce anxiety and guilt or to redirect the child to other activities. However, by age two or three years, most children have internalized previous instances of parental discipline and can mentally represent these rules in the absence of parental instruction. At this age, anxiety in response to transgression is not only concurrent but *anticipatory* (i.e. in response to hypothetical future actions). This anticipatory anxiety inhibits the enactment of future anxiety-causing transgressions.

In the early preschool to kindergarten years (about ages 3 to 6 years), children make significant gains in the ability to recognize different emotions and to make basic inferences about what another person is or might be feeling (Borke, 1971, 1973). Illustratively, Borke (1971, 1973) found that children aged 3 to 6 years demonstrated high accuracy in identifying the emotions of characters in vignettes specially matched to certain emotions. Children of age 3 were 90% successful in identifying "happy" as the emotion experienced by characters in a "happy" story. Rates for correctly identifying "sad" were lower than for "happy" but still significantly higher than predicted by chance, while correct recognition of "fearful" and "angry" was low for younger children but increased across age groups. Cumulatively, children have, by

the school-age years, developed an ability to recognize emotions, to identify situations that might elicit certain emotions for others, to recognize when others might require comfort.

Empathic Skills and The Mechanisms Underlying Their Development

Given the widespread use of low risk, unselected samples in studies of empathic development, the process of socialization described above likely best represents a prototypical process that occurs under relatively ideal circumstances. The dominant theory of development in the field considers empathy development in relation to cognitive development and to morally relevant environmental contexts in which children develop components of conscience, and which may differ substantially between children (Kochanska, 1993). Contingencies in the child's disposition or environment can disrupt this process and result in abnormal or absent empathic functioning. While the literature typically parses these mechanisms roughly into the categories of affective skills (vicarious distress) and cognitive skills (perspective taking), in practice, these skills interact reciprocally to produce mature empathic responses.

In general, empathic internalization occurs as a result of physiological arousal in response to environmental contingencies, including the distress of others and parental prohibition (Kochanska, 1991; Kochanska, 1993). Ideally, the child experiences enough arousal after a transgression that the parental message regarding the transgression is salient and becomes more easily internalized, if not always consciously remembered. Over time, avoidance learning occurs: the child learns to associate various transgressive acts with arousal, which typically takes the aversive forms of anxiety or discomfort (Rose, 1999). The child wishes to avoid distress, so in the future, he or she inhibits the behaviors associated with the distress (Eisenberg, Eggum, & Edwards, 2010). Later in the developmental process, simple anxiety-based avoidance learning is paired with internalized and self-chosen principles that motivate moral actions.

Affective Empathy: The Role of Arousal

Appropriate arousal is perhaps the most fundamental precondition for the development of vicarious distress, which is considered the most basic component of affective empathy. Multiple discussions of empathy impairment suggest that atypical development can occur if the arousal experienced by the child is excessively low or excessively high. These extremes may exist due to congenital dispositions, environmental pressures, or a combination thereof.

Dispositional differences in arousal. Perhaps the best studied example of dispositional effects on socialization is the case of the fearful versus fearless child (Kochanska, 1991, 1993). Fearfulness is a dimension of the higher-order trait of neuroticism that inhibits the potentiation of behaviors (Kochanska & Aksan, 2006). Fearful children avoid exploration in laboratory tasks, while fearless children appear bold and uninhibited (Kagan, 1994; Kagan, Reznick, & Snidman, 1987). Fearful children are also especially likely to experience anxious arousal and to fear re-experiencing that discomfort and incurring consequences (Stifter, Cipriano, Conway, & Kelleher, 2009). Consequently, fearfulness has been linked to appropriately high levels of empathy and moral conduct. Empirical evidence confirms that fearful children are less likely to cheat on tests (Asendorpf & Nunner-Winkler, 1992), experience more empathy/guilt in response to a transgression (Kochanska, Gross, Lin, & Nichols, 2002), and exhibit stronger internalization of rules and standards (Kochanska, 1995; Kochanska, Coy, & Murray, 2001) than fearless children.

In fact, high levels of fearlessness have been associated with impaired acquisition of empathic skills and higher levels of externalizing behaviors (Kochanska, 1991, 1993). This mechanism putatively underlies the origin of poor empathy in children with CU traits, who also demonstrate high fearlessness (e.g. Frick et al., 2014; Frick & White, 2008; Frick, Ray, Thornton, & Kahn, 2014; Frick et al., 1999). Children high on fearlessness are less likely to

experience physiological arousal or empathic arousal in response to distress cues in other people, whether they or another person are the source of the distress (Blair, 1999). If a child experiences too little arousal in response to parental prohibitions, then the prohibition will not be salient to them. They are less likely to attend to and internalize parental messages and are unlikely to be deterred from transgressing, because they fear neither the typical discomfort experienced by others nor the consequences of breaking rules. In support of this notion, high scores on measures of fearlessness predict lower scores on measures of conscience both concurrently (e.g. Kochanska, Gross, Lin, & Nichols, 2002) and longitudinally (Rothbart, Ahadi, Hershey, & Fisher, 2001). In addition, both prospective and cross-sectional work supports a link between low anxiety or physiological arousal and poor emotional empathy, few behavioral expressions of concern towards others, and low levels of guilt following a transgression (Kochanska, 1995; Kochanska, Gross, Lin, & Nichols, 2002; Rothbart, Ahadi, & Hershey, 1994).

While high degrees of fearlessness seem to be maladaptive, too much fearfulness can also be detrimental for moral development in a different way. If the child experiences too much personal distress after a transgression, they may focus more on their own distress than on the parental message, and may consequently avoid transgressions not because they fear harming others, but because they fear their own distress (Eisenberg et al., 2010). In this case, the distress they feel is self-focused rather than other-focused and may encourage withdrawal instead of a reparative or prosocial response. Over time, empathic concern may become blunted or eliminated by the predominant feelings of personal distress.

Cumulatively, the data concerning fearfulness predict an inverted U-shaped curvilinear association between fear and empathy, such that children with moderate levels of fearfulness

demonstrate the highest empathy, while children excessively high or low on fearfulness demonstrate low or impaired empathy.

Environmental modulation of arousal. While temperamental variations in arousal are obvious sources of variation in empathic responding, these variations are likely influenced by the environment as well as innate dispositional differences among children. One of the most obvious sources of environmental moderation operates via parenting. Parenting factors can positively or adversely affect the socialization process. Parental disciplinary styles can disrupt socialization in a manner similar to the previous example of the highly fearful child. The use of excessive or excessively harsh punishment may cause the child to associate their discomfort with the punishment rather than the transgression. Instead of reinforcing the desistance of transgression, this process reinforces the avoidance of punishment. Ultimately, this process conditions the child to feel self-focused personal distress rather than other-focused empathy (Kochanska, 1993). Empirically, authoritarian parenting based on power and harsh discipline is correlated with deficits in the child's later internalization of rules (Kochanska, 1991). Indeed, for highly anxious children, maternal power assertion predicted fewer prosocial responses to a series of vignettes involving hypothetical transgressions (Kochanska, 1991).

Conversely, warm and responsive parenting is associated with more ingrained internalization of rules and empathic concern (Kochanska, 2002). Attachment appears to moderate the effects of fearlessness, such that fearless children who demonstrate secure attachment to their caregivers exhibit less aggression and rule-breaking than children who are insecurely attached (Kochanska, 1995). In this case, children internalize parental messages because they are emotionally close with their parents and desire to please and to emulate them, rather than internalizing prohibitions through traditional avoidance learning. Similarly, a

mutually-responsive orientation (in which parents and children have a relationship based in shared positive affect and cooperation) has been linked to more mature expressions of conscience and more prosocial behavior (Kochanska, 2002).

Given this evidence, ideal socializing discipline is minimally punitive and characterized by warmth and shared positive emotion. Highly fearful children may require gentler discipline, while insufficiently fearful children may require socialization pressure that is stronger but still warm and fair (Dienstbier, 1984). However, exceptions to this model exist empirically. Kochanska (1991) found that for low-fear children, maternal power assertion had few reliable associations with measures of conscience. It is unclear whether this null finding represents a restriction of range (in that effects can only be observed at extremes of low fear) or whether power assertion is only problematic for fearful children who experience too much arousal, while power assertion at levels observed in this study is insufficient to elicit even adequate levels of arousal from low-fear children.

Cognitive Empathy: Perspective Taking

Though arousal is the key piece in the development of affective empathy, it is only one component of a mature empathic response. Deficits in affective empathy seemingly inhibit children from developing cognitive reactions to others' emotions, which impairs the development of cognitive empathy. Over time, these deficits likely operate reciprocally to solidify a pattern of poor empathic responding.

Emotion recognition. The best-studied aspect of emotional processing with respect to empathy is emotion recognition. Emotion recognition is a complex skill that involves an understanding of external features (e.g. facial expressions, situational causes), mental features (e.g. desires, motivations), and cognitive complexities of emotions (e.g. mixed emotions, mental

control of emotions, ability to suppress emotional expression) (Pons, Harris, & de Rosnay, 2004). Emotion recognition underlies the individual's basic ability to feel empathic concern (affective empathy), in that an individual must recognize another person's distress as distress before they can react to it. Recognizing emotions also allows an individual to develop schemas for those situations that elicit certain emotions in other people. This schematic provides the underpinnings for cognitive empathy (perspective taking). Inadequate emotion recognition abilities would putatively impair a child's ability to recognize acts that may distress another person or those that might repair such distress.

In support of this notion, CU traits in children and adults are associated with widely documented deficits in the recognition of and emotional responsivity to fear and potentially sadness (e.g. Blair & Coles, 2000; Blair, Colledge, Murray, Mitchell, 2001; Dadds, Perry, Hawes, et al., 2006; Munoz, 2009; Blair, Budhani, Colledge, Scott, 2005; Fairchild, van Goozen, Calder, et al., 2009). These recognition deficits persist across multiple classes of stimuli, including pictures (Kimonis, Frick, Fazekas, & Loney, 2006), words (Loney, Frick, Clements, Ellis, & Kerlin, 2003), facial expressions (Dadds et al., 2006), and vocal tones (Blair, Budhani, College, & Scott, 2005). On the whole, individuals with severe empathy impairments have difficulty identifying situations in which another person might be afraid or distressed (Marsh & Cardinale, 2012). On the contrary, children who adequately recognize the emotions of others are likely to internalize distress as a victim's potential response and to consider this consequence before acting in an antisocial manner. It is likely that, as these children do not experience the distress of others as aversive or arousing, they do not preferentially attend to fear and sadness stimuli; consequently, they do not adequately learn to identify and distinguish these states (Dadds et al., 2008; Dadds, Perry, et al., 2006).

The Social Environment: Obedience and Compliance

Understanding the development of empathy is important to the extent that empathy predicts observable social behaviors. The degree to which empathy reliably predicts or *must* predict prosocial action, however, is somewhat contested. Empirically, high levels of empathy in children, as assessed by teacher report, have indeed been associated with rule-following behavior (Belacchi & Farina, 2012). However, most research supports the contention that moral emotions and moral behavior only modestly co-occur, particularly in samples of young children (Aksan & Kochanska, 2005; Kochanska, Aksan, & Nichols, 2003; Kochanska, Forman, Aksan, & Dunbar, 2005; Kochanska, Padavich, & Koenig, 1996; Eisenberg et al., 2010; Hartshorne & May, 1928; Zahn-Waxler, Radke-Yarrow, Wagner, & Chapman, 1992). Research has reported stronger, but still modest, associations when situations are analyzed in the aggregate rather than singularly (e.g. Kochanska, Coy, & Murray, 2001). More consistent is the finding that low levels of empathic concern positively correlate with criminal offending, bullying, and aggression in children, adolescents, and adults (e.g. Asendorpf & Nunner-Winkler, 1992; Hastings et al., 2000; Jolliffe & Farrington, 2004; Krettenauer, Asendorpf, & Nunner-Winkler, 2013; Lovett & Sheffield, 2007). While most of this work is correlational in nature, a small number of prospective studies support a causal relationship. For example, Hastings and colleagues (2000) found that children's empathic concern at preschool age predicted lower severity and frequency of externalizing problems at age 6-7 years and lower frequency of externalizing problems at 9-10 years.

Given these findings, recent work by Aksan and Kochanska (2005) and Kochanska and Aksan (2004) suggests that empathy is one of two correlated, but distinct, components of a higher-order construct of conscience. These putative components are moral emotion and moral

conduct (Aksan & Kochanska, 2005; Kochanska & Aksan, 2004). Moral emotion includes the related constructs of guilt and empathy. It is the "motivational engine" that gives transgressions importance to the self and characterizes them as negative and ego-dystonic (e.g. Kochanska & Aksan, 2006). On the other hand, moral conduct exclusively encompasses external behavior and the ability to abide by rules (Kochanska & Aksan, 2006). Each component can exist independently of the other, such that an individual can feel guilt without acting to assuage the suffering of the victim, or can act to aid the victim without feeling congruent distress.

One issue complicating the relationship between empathy and moral responding is the fact that children exist in a social environment that continuously places demands on their behavior. Parents, in particular, are often concerned with empathy and guilt not only because they desire their children to be prosocial citizens, but because they believe that guilt induces compliance with parental commands. Thus, modest correlations may occur due to situational pressures in any given incident, not because moral emotions and moral conduct are linked internally within the child. For example, studies like Milgram's (1963, 1965) suggest that individuals might choose, in a given situation, to obey authority rather than universal principles of morality.

The role of obedience to authority has always been inextricably linked with moral development. Indeed, explicit in Piaget's (1965) and Kohlberg's (Kohlberg & Kramer, 1969) theories of moral development is the dependence of early stages of morality on rules and authority. Even when the authority is not present to give a command, it is implicit in moral decision making that the child is basing their decision to some extent on the proscriptions of a hypothesized or internalized authority figure. A child's decision to act in an immoral or antisocial manner is a function not only of his or her empathy towards a potential victim, but also

of his or her conceptions of the consequences of that action. In this study, our observational assessment of empathy is uniquely able to disentangle these reasons, given that the child must struggle with a choice: obey the adult authority and destroy someone's allegedly cherished item, which could cause harm to the victim, or refuse to destroy the item to protect the victim, thereby disobeying the adult authority.

A History of the Scientific Study of Obedience

The spirit of this dilemma precipitated arguably the most striking study of obedience to authority: that of Stanley Milgram. Milgram's (1963) obedience experiments definitively proved that moral feelings and moral conduct do not universally co-occur when an authority is present and gives commands that conflict with widely held moral principles. In a series of studies, he invited adult males to the laboratory for an experiment purportedly about the effects of punishment on memory (see Milgram, 1963). In each case, the subject was paired with a confederate, and the "random" draw was rigged so that the subject would always be the teacher and the confederate would always be the learner. The subject was presented with a board of buttons with labels ranging from 15 volts "mild shock" to 450 volts "severe shock"; this final button was also labelled "dangerous." The subject was instructed to ask the learner word pairs and to provide increasing shocks if the learner answered incorrectly and was encouraged to continue if he or she protested.

Before conducting the study, Milgram described the experiment to his college class and asked them to guess what percentage of subjects would continue shocking the learner through the last level. This pre-test indicated that students expected between 0 and 3% of subjects (mean = 1.2) would deliver the full amount of shock. These predictions vastly underestimated the actual results. Of 40 participants, none discontinued the experiment before reaching a shock of 300

volts. Five subjects refused to continue beyond 300 volts and 14 subjects desisted before the final level, leaving 26 participants who obeyed the experimenter to the final shock level.

While this first experiment provided evidence *that* a subject would obey, it did not empirically investigate *when* the subject would obey, that is, what circumstances (internal or external) moderate compliance. In a series of adjustments to the original paradigm, Milgram investigated a variety of factors that moderated obedience (Blass, 1991). Proximity was a powerful factor; the more salient and visible the victim, the less likely the subject would obey until the final shock. Thus, many subjects complied when the learner was in a different room or provided verbal feedback, but fewer complied when the learner was in the same room and even fewer complied when they were required to deliver the shock by physically pushing the learner's hand onto a shock plate (Milgram, 1965).

Status and salience of the authority figure also moderated compliance. Orders delivered over the phone decreased compliance (Milgram, 1974). Compliance also decreased when another "subject" (a confederate) gave the orders. This is consistent with work by Shalala (1974) in a military setting; subjects were more likely to comply with a higher ranking officer than a lower ranking officer. Finally, the presence of others altered compliance. Fewer participants complied when a second teacher (a confederate) was present and desisted, and when two experimenters were present and gave conflicting orders (stop vs. continue) (Milgram, 1965, 1974).

These early studies gave rise to the notion of "power of the situation" (Gaertner, 1976; Ross, 1977; Zimbardo, 1974). This concept describes the finding that the demand characteristics of certain situations constrict behavioral responses more than others—the idea of "strong" versus "weak" situations. Strong situations tend to suppress the effects of dispositional characteristics;

these are situations in which no choices are given and may be imposed rather than freely chosen (Benjamin & Simpson, 2009; Blass, 1991). Indeed, such emphasis on the power of the situation led many researchers to conclude that research into individual differences was untenable, and that situational factors, not putative personality traits, accounted wholly for human behavior (see Epstein & O'Brien, 1985 for a historical review of the person-situation debate).

Yet, obedience is not wholly environmentally motivated. Despite the high rate of obedience in Milgram's (1965) seminal study, a number of participants did not obey the experimenter to the conclusion of the study; additionally, rates of obedience dropped when additional factors were introduced (Blass, 1991). In subsequent research, many personality traits have been associated with obedience. In weaker situations, in which choices exist and behaviors may be freely chosen, dispositional characteristics have been shown to have stronger effects on obedience than in strong situations (Benjamin & Simpson, 2009; Snyder & Ickes, 1985). Across studies, researchers have found that greater compliance with authority is linked to individual differences in authoritarianism (Elms & Milgram, 1966), trust that the experimenter is benign and would not do real harm (Mixon, 1976), lower stages of moral development according to Kohlberg's stages (Milgram, 1974), lower social intelligence (Burley & McGuinness, 1977), and low suspiciousness (Holland, 1967). Dispositional aggression has been found to relate to obedience in studies modeled after Milgram's shock paradigm when a choice is given (e.g. the subject may choose which level of shock to give) rather than strict limitations (e.g. subject must give increasing levels of shock or disobey). Despite evidence supporting strong situations as powerful determinants of behavior, a multitude of research still supports the importance of individual differences in predicting human action.

Obedience in Children

The role of obedience in the moral actions of children presents an especially complex case. Whereas adults are presumed to judge moral situations based on a variety of nuanced factors, strict obedience to authority has historically been considered the primary motivation for moral behavior in young children. Indeed, Piaget (1965) and Kohlberg (Kohlberg & Hersh, 1977) posited that young children regard rules as obligatory, moral decisions as functions of reward or punishment, and adults as unilaterally legitimate authority figures to be obeyed without question (Krahn, 1971).

Yet, research on compliance in children rejects the idea that children obey authority figures unquestionably or consider adults to be always legitimate moral authorities. In fact, an accumulation of evidence suggests that children view authority and their obligation to obey (or defy) in nuanced ways, and are not simply passive recipients of adult commands (Kuczynski, Kochanska, Radke-Yarrow, & Girnius-Brown, 1987). Nor is all compliance equal in degree. Kochanska & Aksan (1995; see also Kochanska, 2002) distinguish between committed and situational compliance. Committed compliance exists when the child complies wholeheartedly without prompting or reminding, and appears to embrace the maternal (or parental) agenda. Situational compliance involves the child following directions, but requiring assistance or reminders to stay on task, seemingly because he or she does not fully embrace the maternal agenda.

Kochanska & Aksan (1995) posit that these types of compliance may reflect different motivations for complying, and thus different levels of moral internalization. Across studies, committed compliance predicts moral internalization, both concurrently and prospectively (e.g. Forman & Kochanska, 2001; Kochanska & Aksan, 1995; Kochanska et al., 1995; Kochanska,

Coy, & Murray, 2001; Kochanska, Tjebkes, & Forman, 1998). On the other hand, situational compliance is externally regulated and not associated with internalization (Kochanska, 2002).

Qualitative research has illuminated the factors that determine whether children will comply with an authority. Empirically, children have cited that they comply with commands in order to avoid punishment, to obtain rewards, to please their parent, peer, or other authority, or to serve the needs of others (Carlsmith, Lepper, & Landauer, 1974; Laupa, 1991; Lundy, Shell, & Roth, 1985). Children also demonstrate relatively sophisticated notions of authority (Damon, 1977; Laupa & Turiel, 1986). Even children as young as four years old judge whether they should obey a command based on the legitimacy of the authority. Laupa & Turiel (1986) and Laupa (1991) found that younger children (age 6-8) expressed that they were more likely to obey an adult (teacher) than a peer granted authority by the school, while older children (9-10 years old) were equally likely to see the teacher and peer authority as legitimate.

Additionally, children determine whether they should obey based on the type of act commanded. Importantly, the compliance literature based on observed behavior is predicated almost exclusively on the child's obedience to benign commands (i.e., pick up the toys, don't touch the toy), while the literature concerning obedience to morally relevant commands (i.e., steal the toy, fight with the child) is almost exclusively predicated on hypothetical vignettes. Vignette studies have universally found that most children are likely to judge a command to clean their room or pick up toys as legitimate and obligatory, while they view commands to steal or fight as illegitimate and not to be obeyed, even if the command comes from a parent or teacher (Damon, 1977; Laupa, 1991, 1994). Most children even find the commands of a nonauthority more legitimate than the commands of an authority figure if the authority figure is commanding a harmful act (Laupa, 1994). However, this pattern was strongest for older

children; some younger preschoolers still expressed that they should obey the adult authority's command to fight with another student because the command was given by an adult (Laupa, 1994). Given that behaving morally is socially desirable, it is possible that the demand characteristics of vignette studies pull children to supply what they perceive as the "right" (or morally superior) response to the adult experimenter, even if their actual behavior would differ.

A very small body of literature has investigated whether children will obey immoral commands in real situations, even if they disavow such actions when presented with a hypothetical. In a striking example, Shanab & Yahya (1977) replicated Milgram's shock paradigm in a group of children aged 6 to 15. Overall, 73% of the subjects continued to deliver shocks until the final level, a rate strikingly similar to the results reported in Milgram (1965). Across ages, girls were more likely to display signs of tension and were more likely to state that they complied because they were obeying orders; boys were more likely to cite that fear of punishment is beneficial for learning when asked why they complied.

Finally, temperamental factors also play a role in compliance. Children who demonstrate high negative emotional reactivity exhibit more noncompliance than their peers, although this finding could also be a function of the fact that these children also elicit greater maternal control, which is itself related to noncompliance (Braungart-Rieker, Garwood, & Stifter, 1997). Anxiety appears to have an interesting role in compliance. Carlsmith, Lepper, & Landauer (1974) found that the motivation for compliance seemed to change based on anxiety. Children induced to be anxious were more obedient to an experimenter previously experienced as negative, perhaps to avoid disapproval. On the other hand, children not induced to feel anxious seemed to be more reward-focused, and were more obedient to a positive experimenter. While induced anxiety is not identical to temperamental anxiety, this study suggests that the experience of anxiety during

a given situation affects the child's perception of that situation. In another study, low anger and low social fearfulness was also related to compliance (Kochanska, 1993).

The predictive power of some temperamental factors may differ depending on the type of task (Kochanska, 1993; Kochanska, Coy, & Murray, 2001; Lickenbrock, Braungart-Rieker, Ekas, et al., 2013; Rothbart & Bates, 1998). Compliance has typically been measured by parent report and by observational tasks. These tasks fall into two categories: "Do" tasks, in which the child is asked to perform a behavior, and "Don't" tasks, in which the child is asked *not* to perform a behavior. Across the literature, a clean-up paradigm is typically used for "Do" tasks and a forbidden toy paradigm is used for "Don't" tasks. Incidentally, most studies have found that "Do" tasks elicit less committed compliance and more situational compliance or noncompliance than "Don't" tasks (Braungart-Rieker, Garwood, & Stifter, 1997; Kochanska, Aksan, & Koenig, 1995; Kochanska, Coy, & Murray, 2001). In fact, committed compliance typically demonstrates weak or nonsignificant correlations across task type.

Researchers suggest that each type of task acts on different aspects of a child's temperament, thus contributing to the lack of consistency in compliance observed across task types (Braungart-Rieker, Garwood, & Stifter, 1997). One important aspect is self-regulation. Compliance can be conceptualized as an early form of self-regulation, in that it requires the child to modulate his or her behavior in accordance with situational demands (Kochanska, Coy, & Murray, 2001). However, effortful control, a form of self-regulation, appears to preferentially predict compliance in "Do" tasks, with only modest associations with "Don't" tasks (Kochanska, Coy, & Murray, 2001; Kochanska et al., 1997). Fearfulness, on the other hand, has been found to predict compliance in "Don't" tasks but not in "Do" tasks (Kochanska, 2002; Kochanska, Coy, & Murray, 2001). This is consistent with psychopathy research, which finds

that individuals high on fearlessness have difficulty suppressing prohibited acts (Kochanska, Coy, & Murray, 2001). However, the content of both "Do" and "Don't" tasks almost exclusively involves benign scenarios (e.g., pick up your toys, don't touch the forbidden toy) rather than commands to misbehave or cause harm.

Importantly, the vast majority of research on child compliance has operationalized compliance as obedience to a parent, usually the child's mother. Of studies that consider obedience to another authority, most utilize child report measures or responses to hypothetical vignettes. The few studies that have investigated this question observationally have found relatively high levels of obedience to non-parent authorities (Landauer, Carlsmith, & Lepper, 1970; Shanab & Yahya, 1977). In fact, Landauer, Carlsmith, & Lepper (1970) had children receive commands from three separate adult women, one of whom was the child's mother. Overwhelmingly, children complied more with commands from the unrelated adult females than with commands from their own mothers. Children's compliance to their own mothers' requests is likely a function, in part, of the quality of the parent-child relationship (Abe & Izard, 1999; Crockenberg & Litman, 1990; Kochanska & Aksan, 1995; Kuczynski & Kochanska, 1990), while situational or temperamental factors may be more important when dealing with a novel adult.

Finally, compliance also depends in part on the tactics used to elicit compliance. Maternal control is a particularly robust predictor of compliance; in fact, Kuczynski & Kochanska (1990) found that the best predictor of child compliance assessed at age 5 was the type of maternal control assessed at toddlerhood, with power assertive tactics predicting the poorest compliance. Negative types of control have been associated with defiance (e.g. saying "no" with a degree of oppositionality) and passive noncompliance (e.g. ignoring the command)

(Abe & Izard, 1999; Crockenberg & Litman, 1990). Maternal control also preferentially predicts situational, as opposed to committed, compliance (Braungart-Rieker, Garwood, & Stifter, 1997). On the other hand, warm maternal guidance has been associated with assertive and compliant behavior (Kochanska & Aksan, 1995).

Changes in Compliance Across Development

Irrespective of other factors influencing compliance, the bulk of research has found that compliance increases as children age. Most studies report that older children demonstrate more committed compliance and less noncompliance than younger children, both in studies comparing cohorts and in prospective studies (Braungart-Rieker, Garwood, & Stifter, 1997; Kochanska & Aksan, 1995; Kochanska, Aksan, & Koenig, 1995). Many studies also report higher levels of compliance for girls than for boys (Kochanska, 2002; Kochanska, Coy, & Murray, 2001; Kochanska, Tjebkes, & Forman, 1998). Additionally, Kochanska (2002) found specifically that compliance in girls was related to self-reported view of a moral self and internalization, while there was no correlation between these constructs and compliance for boys. Perhaps boys' compliance is more related to external situational characteristics.

Forms of noncompliance also demonstrate differing frequency across development. Passive noncompliance, considered a less mature form of noncompliance, decreases across toddlerhood, while active opposition, including both assertiveness and oppositionality, increases (Kuczynski, Kochanska, Radke-Yarrow, & Girnius-Brown, 1987). However, uncontrolled anger and defiance tends to decrease by age 7 (Kuczynski, Kochanska, Radke-Yarrow, & Girnius-Brown, 1987). Thus, as children age, not only are they better able to comply, but the tactics by which they refuse to comply become more sophisticated and less emotionally dysregulated.

The Present Study

Despite significant advances in the field of moral and empathic development, the understanding of moral decision making in real-life situations is limited. Historically, the moral development literature in children has been hampered by methodological difficulties. Empathy is typically operationalized through self-report (e.g. De Wied, van Boxtel, Matthys, Meeus, 2011) and parent or teacher report (e.g. Dadds et al., 2009). Even parents reporting on their child's CU traits must make inferences about their child's internal motivations for behavior (i.e., callousness versus environmental provocations). Additionally, it is socially desirable to be empathic, which suggests that parents may be motivated to rate their children as being more empathic than they are. While parent report provides invaluable data, parents are likely unaware of the extent of their child's moral decision making; they can only infer their children's moral rationale from their child's behaviors.

In attempt to circumvent this issue, the majority of studies of moral development have utilized hypothetical, morally charged vignettes in order to assess the child's own moral decision making (e.g. Damon, 1977; Laupa, 1991, 1994). These scenarios typically involve a situation in which the child must make a moral choice; for example, the child may be asked what he or she would do if a peer picked a fight, and why he or she would make that decision. Vignette studies have provided invaluable evidence about how children navigate the moral world, including how they incorporate authority-based rules, social norms, personal empathic feelings, and broader moral principles into their decision making. However, there is little research to determine whether children's answers to hypothetical scenarios correlate with their actual behavior in dayto-day life.

A minority of studies has empirically examined moral decision making through performance-based laboratory tasks (Brook & Kosson, 2013; Shanab & Yahya, 1977). Performance-based tasks may reduce some of the problems with questionnaire data, particularly for young children. Although, as with self-report, these tasks require some inferences about internal states, studies have found compelling predictive associations between laboratory task behavior and aspects of moral development. For example, Kochanska, Murray, Jacques, Koenig, & Vandegeest (1996) found that effortful control as assessed via a forbidden toy paradigm was correlated with children's internalization of moral principles.

Thus, laboratory tasks can circumvent the problematic lack of consistency between questionnaire measures of a child's past or hypothetical behavior and their actual behavior in a morally relevant scenario. The aforementioned methodological issues suggest that a deeper understanding of empathy, moral decision making, and its relation to externalizing problems should include a greater emphasis on observational data, and should attempt to disentangle the multiple factors at play.

In the present study, we used just such a laboratory observational approach to explore the nature of children's affective, social, and behavioral responses to a morally rich scenario, collected during the course of a multi-task laboratory assessment of child temperament. This study explored individual differences in these responses and their associations with child age, sex, temperament traits, verbal intelligence, and emotion recognition abilities. Further, because the moral task was collected in the context of a larger study assessing the development of temperament and risk for psychopathology with multiple assessments collected over 2 years, this study also explored concurrent and predictive associations between children's responses to the moral scenario and the development of externalizing psychopathology in particular.
Aims of the Present Study

Overarching Aim: Validate the picture tearing coding scheme as an assessment of child moral development characteristics. For many reasons already enumerated, the observational Picture Tearing task provides a rich and unique opportunity to assess aspects of child moral development. However, while this task is potentially illuminating and is described as part of the Laboratory Temperamental Assessment Battery (Goldsmith et al., 1995), no published study to date has included this task. Moreover, the coding scheme suggested in the original Lab-TAB manual (Goldsmith et al., 1995) has not been validated in a published paper. Our coding scheme is based on the original scheme, but includes additional variables of interest to moral development and increases variability by substituting Likert scales for previously dichotomous items.

In order to support this new scheme as an effective tool for future research, we have examined its utility in describing and predicting behaviors of interest. Specifically, we have determined to what extent this coding scheme predicts behaviors both within the task and outside of the lab (externalizing behaviors, in particular). We have used the results of our analyses to make recommendations as to which variables are most valid for tapping individual differences in moral development and which have the greatest utility for predicting outcomes related to moral development (i.e., externalizing problems). Thus, the present study may help to encourage use of laboratory based observational methods for assessing moral development, and illuminate the specific codeable behaviors most valid for understanding aspects of moral development and predicting conceptually related outcomes of applied interest. In order to achieve this overarching goal, we divided our analyses into three specific aims that each address an issue important to assessing the overall utility and validity of our instrument.

Specific Subordinate Aims

Aim 1: Describe variations in empathic responding in a morally relevant scenario. To date, no study has simultaneously addressed the associations of empathy, guilt, and obedience with morally relevant behavior via observational methods in youngsters. The present study aimed to address gaps in the moral development literature by assessing the characteristics of empathy, obedience, and noncompliance in young children in a naturalistic moral scenario: the Picture Tearing task (Goldsmith et al., 1995). In this task, the child is asked by one adult experimenter (with whom s/he has been interacting for over an hour) to perform an immoral act (tearing the other experimenter's favorite picture). The second experimenter then returns to observe what the child has done (or not) to the allegedly cherished object. This task is specifically designed to elicit moral emotions (e.g. empathy, guilt) and to pit higher levels of moral reasoning (e.g. appeals to universal moral principles) against more rudimentary levels (e.g. obedience to authority). The children in our study ranged from 3 to 7 years old; across this age range, children should demonstrate an ability to recognize emotions, to identify situations that might elicit certain emotions for others, to recognize when others might require comfort. Thus, most of the children, especially in the older half of the age range, should theoretically be able to take the perspective of the second experimenter and identify that she might feel upset if her favorite picture is destroyed. However, we anticipated there would be wide individual differences in all aspects of responses to this task.

Ethical concerns prohibit the use of shock paradigms as in prior obedience research (Milgram, 1963; Shanab & Yahya, 1977). Thus, as opposed to physical harm (a shock), the children in our study were asked to cause emotional harm. This type of immoral act is subtler, but also more similar to the kinds of dilemmas that children typically encounter. Additionally,

this paradigm allowed us to assess moral emotions and moral reasoning at multiple levels. First, we could observe the child's actual choice to obey an authority figure and to perform or refuse an immoral action. Second, we could observe behavioral indicators of empathy and guilt, including verbalizations, both during the decision-making period and when the child is confronted by the person they have wronged (for those who choose to tear the picture). These data will be used to describe patterns of empathic responding and obedience among children, including descriptive data on rates of compliance with the Picture Tearing Request. We expected to find rates of obedience that are comparable with or higher than the rates reported in Milgram (1963) and Shanab & Yahya (1977), given that children tend to appeal more to authority (e.g., the primary experimenter in this study) when making such decisions (compared to adults). We also examined whether the gender of the second experimenter (the "victim") affected children's responses to the request.

Aim 2: Examine associations between observed moral behavior and child characteristics theoretically linked to moral development and empathy. Not only could we observe the child's moral decision making in the moment, but we could also test associations between the child's behavior in response to the Picture Tearing task and other dimensions that have been linked empirically to moral decision making in the literature, including normative factors of sex, age, temperament, emotion recognition ability, and intelligence. Foremost, we wanted to establish the construct validity of the Picture Tearing task as a measure of individual differences in aspects of moral development. If the task produces a valid measure, then scores on Picture Tearing composites should correlate with known correlates of moral development according to patterns observed in previous research. We tested the relevance of empathy measures assessed in this task for understanding abnormal development related to deficits in

empathy, specifically risk for CU traits and related problems (assessed both concurrently and predictively). In this way, we clarified the temperamental characteristics associated with variously sophisticated levels of empathic responding, and also investigated the behavioral factors observed in this task that are predictive of later aggression and delinquency. We also explored broader contextual contributors to moral development and empathy, specifically aspects of the parent-child relationship. Finally, we investigated the incremental validity of each of these domains for understanding child behaviors in response to the Picture Tearing task, including compliance, guilt, and empathy, via hierarchical regression. Analyses are described in more detail below.

Aim 3: Examine predictive associations between compliance, arousal, and empathy and later externalizing behaviors. Finally, we examined whether observationally assessed compliance, arousal, and empathy at baseline will predict outcomes at 6 months, 9 months, 12 months, 18 months, and 24 months after the initial assessment. We were especially interested in the prediction of externalizing pathology, given its robust empirical associations with low empathy, poorly developed conscience, and disorders of empathy (i.e., CU traits) (e.g. Asendorpf & Nunner-Winkler, 1992; de Wied, van Boxtel, Matthys, and Meeus, 2011; Frick et al., 2003; Frick & Morris, 2004; Frick, Ray, Thornton, & Kahn, 2014; Frick & White, 2008; Jolliffe & Farrington, 2004). Given the accumulation of evidence, we predicted that empathy and arousal during the Picture Tearing task would predict externalizing problems (e.g. CU traits, aggression, rule-breaking, ODD symptoms, CD symptoms) at all follow-up time points, over and above the variance associated with baseline levels of these behaviors. To assess discriminant validity, we also examined the relationship between empathy and arousal and later internalizing problems.

METHOD

Overview of Design

A community sample of child participants (N = 271 children from 220 families) aged 3 to 7-years-old and their parents was recruited from the central Michigan area for a study of child temperament. Participants were recruited through advertisements on the Cragslist website as well as local postings in areas frequented by children (e.g. preschools). Children who did not have any significant medical conditions or developmental disabilities and lived with at least one English-speaking biological parent were eligible for participation in the study. Participating children visited the laboratory with their mother or father for a 2-hour assessment consisting of 16 tasks designed to elicit discrete emotions and behaviors indicative of temperament traits, as well as assessments of intelligence and emotion recognition skills. At the end of the lab visit, the parent was given a battery of questionnaires assessing temperament traits, problem behaviors, and parent-child relationship to complete and return by mail.

Follow-up packets of questionnaires assessing child temperament traits and problem behaviors were mailed to participants at 6 months, 9 months, 12 months, 18 months, and 24 months after the initial visit. These packets were to be completed at receipt and returned by mail. When the date of questionnaire completion was ambiguous (i.e. the participants mailed back a packet several months after having received it), graduate assistants contacted the family to obtain an accurate date. If the family failed to return three follow-up packets in a row, they were considered uninterested in further assessment and removed from the mailing list. Participants were compensated monetarily (gift cards) for the baseline assessment and for each returned follow-up packet.

Participants

Child participants were between the ages of 3 and 7 years at baseline (49.4% girls). The mean age of the children was 54.5 months (SD = 16.3). Data on race and ethnicity and family income were provided by 64.6% of mothers and by 40.4% of fathers. Of those, the ethnic composition of mothers was as follows: Caucasian/White (76.5%), Hispanic/Latino (8.4%), African American/Black (5.0%), Asian (1.7%), Native American (1.1%), other (1.7%), and bi- or multiracial (5.6%). The ethnic composition of fathers was as follows: Caucasian/White (73.2%), Hispanic/Latino (6.3%), African American/Black (10.7%), Asian (0.9%), Native American (2.7%), other (1.8%), and bi- or multiracial (4.5%). Participants were not asked about the ethnicity of their child. In all analyses where ethnicity is used, we used mother ethnicity as a proxy for child ethnicity, as almost all children had their mother participating, while father participation rates were substantially lower. In the few cases where the child's mother did not participate, we instead used the father's ethnicity. Yearly family income ranged from less than \$10,000 to greater than \$100,000. Income distribution was as follows: under \$10,000 (5.8%), \$10,000 to \$20,000 (15.1%), \$21,000 to \$40,000 (21.5%), \$41,000 to \$60,000 (30.8%), \$61,000 to \$100,000 (22.1%), over \$100,000 (4.7%).

Laboratory Assessment of Temperament

Children completed a battery of 16 laboratory tasks designed to assess differences in three broad dimensions of temperament: positive emotionality (PE), negative emotionality (NE), and effortful control (EC). The battery was composed of episodes from the Laboratory Temperament Assessment Battery—Preschool Version (Lab-TAB; Goldsmith, Reilly, Lemery, Longley, & Prescott, 1995) and from previous investigations (e.g. Durbin, 2010; Durbin et al., 2007; Frye, Zelazo, & Palfai, 1995; Kochanska et al., 1996; Zelazo, Muller, Frye, Marcovitch,

2003). Tasks were administered in the same order across participants. Parents were present for the tasks as noted. For these tasks, parents were instructed to remain affectively neutral and to minimize their interaction with their child. Tasks are described in full below.

Exploring New Objects (Durbin, 2010; Fear, happiness). The child was left to explore the room, which contained novel and ambiguous stimuli including: a tunnel containing a mechanical spider, an animal crate containing toy mice, a wooden box containing sticky "worms," and a plastic skull hidden under a cloth. The experimenter returned after 4.5 minutes and asked the child to interact with each object.

Making a t-shirt (Durbin, 2010; engagement, happiness). The child decorated a t-shirt with ink and stamps. The child was allowed to take the t-shirt home at the end of the lab visit.

Stranger approach (Lab-TAB; Goldsmith et al., 1993; Fear). The child was left alone briefly in the testing room. A male research assistant entered the room and engaged in a scripted conversation with the child using a neutral voice. The experimenter then returned and introduced the research assistant as a friend.

Impossibly perfect green circles (Lab-TAB; Goldsmith et al., 1993; Anger, sadness). The experimenter repeatedly asked the child to draw green circles on a piece of paper while mildly criticizing each circle. After 2 minutes, the experimenter positively commented on the child's circles.

Popping bubbles (Lab-TAB; Goldsmith et al., 1993; Activity, happiness). The experimenter made bubbles with a bubble-shooting toy and encouraged the child to pop the bubbles.

Diorama snakes (Fear). The experimenter showed the child a tray filled with sand and two remote-controlled snakes and asked the child to touch the snakes. A second experimenter

(holding the remote controls out of sight of the child) made the snakes move. Following the child's response (e.g. touching the snake), the experimenter explained that the snakes are toys and demonstrated the use of the remote controls.

Snack delay (Lab-TAB; Goldsmith et al., 1993; Effortful control). The experimenter placed an M&M or other small snack under a clear cup and instructed the child that he/she could eat the M&M when the experimenter rang a bell. The experimenter rang the bell at eight predetermined time intervals (ranging from immediately to 30 seconds after M&M placement).

Picture tearing (Goldsmith et al., 1995; Guilt, empathy, sadness, compliance). The child's parent was not present for this task. The main experimenter told the child that she will prepare the next game while the second experimenter shares pictures with the child. The second experimenter showed the child a series of pictures depicting a fictional vacation home and grandparents. The second experimenter noted that the picture of the grandparents is his/her favorite because he/she rarely sees his/her grandparents. The second experimenter left the room. The main experimenter returned and asked the child to tear up the second experimenter's favorite picture. This command was repeated several times. After the child's response, the main experimenter excited the room. The second experimenter then returned to the room to retriever his/her photo album. If the child tore the picture, the second experimenter stated, "Oh no! What happened?" and allowed time for the child to respond. The second experimenter then retrieved a second copy of the picture and had the child help return it to the album. The main experimenter then emphasized that she should not have asked the child to tear the picture. Intraclass correlation coefficients are presented in Table 1. ICCs for most variables was at 0.70 or better, though a small number of variables with very low base rates (e.g. Hunched Shoulders, Lip

Biting) had lower ICCs. This was taken into account when selecting variables to retain for analysis.

Balloon bop (Goldsmith & Rothbart, 1996; Effortful control, happiness). The experimenter and child played a game in which they took turns hitting a balloon in the air. The child was told that he/she must keep their feet within a circle drawn on the ground. The experimenter tempted the child to leave the circle by hitting the balloon further from the circle on some trials.

Transparent box (Lab-TAB; Goldsmith et al., 1993; Anger, sadness). The experimenter locked an appealing toy chosen by the child in a clear plastic box and gave the child an incorrect set of keys. The experimenter told the child to open the box with the keys, and then left from the room. After three minutes, the experimenter returned with the correct set of keys and explained that she accidentally gave the child the wrong keys. The child was then allowed to open the box and play with the toy.

Simon says (Strommen, 1973; Effortful control). The experimenter demonstrated ten exercises (e.g., touch the floor) to the child and had them practice. The experimenter then told the child that in the game, the child should do all of the exercises when the experimenter says, "Simon says," and should not do the exercises when the experimenter does not say, "Simon says." After two practice trials, the experimenter completed two trials each of all 10 exercises, regardless of the child's performance.

Tell a story (Durbin et al., 2007; Fear). The experimenter told the child that she would like to know how good the child is at telling stories. The experimenter handed the child a picture book without words and told the child to tell a story using the pictures. The experimenter explained that a second experimenter is an expert at stories and will listen to the child's story and

give him/her a grade. At the end of the story, the experimenter asked the second experimenter to give the child a grade on the story.

Pop-up snakes (Lab-TAB; Goldsmith et al., 1993; Anticipatory positive affect,

happiness, surprise). The parent was not present for the first half of the task. The experimenter showed the child what appeared to be a can of potato chips, which instead contained coiled spring snakes. The experimenter demonstrated the trick and then encouraged the child to play the trick on his/her parent.

Walk-a-line slowly (Kochanska et al., 1996; Effortful control). The experimenter demonstrated how to walk a line on the floor and asked the child to walk on the line. The experimenter then asked the child to walk the line as slowly as he/she could, then as quickly as he/she could, then again as slowly as he/she could. The experimenter then asked the child to walk as slowly as he/she could on a small balance beam.

Box empty (Lab-TAB; Goldsmith et al., 1993; Anticipatory positive affect, anger,

sadness). The experimenter gave the child an empty gift bag under the pretense that the bag contains an appealing gift. The child was told to open the gift and was left alone for 2.5 minutes. The experimenter then returned with a bag of toys and told the child that she forgot to put the toys inside the gift bag.

Coding of Child Temperament Laboratory Tasks

Observational coding of picture tearing task. Picture tearing videos were coded with a global system (see Appendix A). Codes were derived from the original coding system developed for the task by Goldsmith and colleagues (2001). Items retained were: Latency to tear up picture (in seconds), Presence of enjoyment in tearing picture, Overall compliance, and Relief in replacement picture. The ratings for Presence of enjoyment, Overall compliance, and Relief in

replacement picture were expanded to four-point Likert scales in order to increase the possible range of scores (variance) on these dimensions. Ratings for intensity of sadness, anger, empathic statements, and visual referencing were retained, but were coded as counts (i.e., the number of discrete instances during which these expressions occurred) rather than global dimensions. In addition, counts were added to assess additional behaviors including: expressions of defiance, expressions of hesitation, squirming/fidgeting, and gaze avoidance (presence of these behaviors was coded every five seconds).

Behaviors were counted and summed separately across three epochs: during the prompting phase (between the experimenter's first prompt to tear the picture and the experimenter leaving the room), during the period in which the child is left alone in the room, and during the period after the second experimenter returned and asked what happened to the picture. Specifically during the third epoch, counts were taken for the child's responses to the experimenter's inquiry, which included: blaming someone else, blaming themselves, silence, or lying. A global code for guilt ("How guilty did the child seem?") was rated on a four-point scale. For this code, coders were instructed to rate the child's overall guilt based on a global evaluation of their behavior, taking into account an integrated impression of facial expressions, posture, and verbalizations.

Figure 2 and Table 2 summarize the codes relevant to each epoch.

Observational coding of laboratory tasks. All laboratory episodes (including Picture Tearing) were coded with a global system validated in previous studies of temperament to assess higher-and lower-order facets of Positive Emotionality, Negative Emotionality, and Effortful Control. Of particular importance for this study, lower-order traits of fear proneness and compliance were rated in all tasks. This enables us to explore whether individual differences in

children's anxious arousal/fear or compliance with the experimenter in the Picture Tearing task are correlated with the same traits assessed in other lab scenarios that lack the moral dilemma of Picture Tearing and whether associations between fear and compliance in response to Picture Tearing are associated with outcomes of interest after accounting for general trait fear proneness and trait compliance. Coders were trained undergraduate and graduate students.

Coders recorded instances of facial, vocal, and bodily indicators of emotional states (i.e. happiness, sadness, fear, anger, surprise). Indicators will be rated for intensity according to the AFFEX coding system (Izard, Dougherty, & Hebree, 1983), which classifies intensity of facial expressions in three levels: (1) ambiguous or low intensity (expressions of low intensity in one facial region); (2) moderate intensity (expression definitely present in at least one facial region); and (3) high intensity (expression definitely present in both facial regions, i.e. eyes and mouth). Intensity of vocal and bodily expressions were also coded on a three-point scale (low, moderate, high) as determined by the extent to which the vocalization or bodily movement conveyed the emotion.

Coders also completed global ratings of the child's behavior during each laboratory task. We will utilize the scales indexing the child's level of Compliance, Engagement in the task, Initiative, Sociability, Attentional Control, and Impulsivity. Each of these behaviors was rated on a four-point Likert scale (0 = low, 1 = moderate, 2 = moderate-to-high, 3 = high). Each behavior code is assigned a single rating (0-3) based on the aggregate of behaviors for a given episode. Compliance ratings are based on the extent to which the child acts in accordance with the parent's or experimenter's suggestions or commands. Engagement ratings are based on the child's degree of attentiveness to and persistence throughout the task. Initiative ratings are based on the child's degree of assertiveness in interactions with the experimenter or parent. Sociability

ratings are based on the child's interest in and pursuit of social interaction with the experimenter or parent. Attentional Control ratings are based on the child's ability to maintain attention flexibly across the task. Impulsivity ratings are based on the child's tendency towards impatient and impulsive behavior, i.e. a lack of appropriate and planful behavioral control.

Codes were aggregated across episodes as appropriate, using scale construction techniques (evaluation of internal consistency reliability, average inter-task correlations, and interrater reliability). Aggregate codes will be combined into higher order dimensions for traits of Positive Emotionality (PE), Negative Emotionality (NE), and Effortful Control (EC), consistent with evidence from other studies using laboratory assessment of traits (e.g. Durbin, Hayden, Klein, & Olino, 2007; Dyson et al., 2012; Hayden, Klein, Durbin, & Olino, 2006; Vroman, Lo, & Durbin, 2014; Wilson & Durbin, 2012). In this system, the codes that combine to create each higher-order code are as follows: Negative Emotionality is comprised of Fear, Sadness, and Anger; Positive Emotionality is comprised of total Positive Emotionality, Anticipatory Positive Emotionality, Sociability, Interest/Engagement, Initiative, and Activity; Effortful Control is comprised of Noncompliance, Attentional Control, and Impulsivity.

Other Tasks

Emotion recognition training. Children were shown six cards with pictures of a boy or girl demonstrating the following facial expressions: Fear, Surprise, Neutral, Sad, Happy, Angry. The experimenter pointed to each picture in turn and asked the child to describe how the person in the picture was feeling. Children's responses were recorded and a total score was given for the number of correct labels the child provided. Children were given feedback if they were correct and were provided with the correct label if they were incorrect. The cards were then shuffled and the experimenter asked the child to identify all six emotions again. Feedback was

given for any incorrect emotions. If the child correctly labeled all six emotions, the task was discontinued. If the child labeled any emotion incorrectly, a third trial was performed, and the final performance score (i.e., number of emotions correctly labeled on the third trial) was recorded. Emotion recognition scores were available for 97% of children. Overall accuracy on the first and final trial administered were recorded and were used to assess basic emotion recognition skills. We calculated scores (total correct) separately for the first trial, which assesses the child's baseline knowledge, and the third trial, which assesses their knowledge after training. We examined each of these variables separately in analyses. Deficits in emotion recognition, especially for fear and sadness, are empirically related to deficits in empathy and guilt (e.g. Blair & Coles, 2000; Blair, Colledge, Murray, Mitchell, 2001; Dadds, Perry, Hawes, et al., 2006; Munoz, 2009; Blair, Budhani, Colledge, Scott, 2005; Fairchild, van Goozen, Calder, et al., 2009).

Peabody Picture Vocabulary Test (PPVT). Children were administered the Peabody Picture Vocabulary Test (PPVT; Dunn & Dunn, 1997) to assess global cognitive function. The PPVT is a widely used measure of receptive vocabulary consisting of 204 items. For each item, respondents are presented with four pictures and are asked to identify which picture best corresponds to a given word. Respondents begin with the set of items corresponding to their age; if the respondent makes fewer than 2 errors, that set is considered basal. Items are administered until the respondent makes 8 or more errors in any given set. Standardized PPVT scores were utilized in all analyses. PPVT scores were available for 99% of children.

Measures

Child Behavior Checklist (CBCL). Participating mothers and fathers completed the Child Behavior Checklist (CBCL; Achenbach & Rescorla, 2001), which was designed to

measure behavior problems in children. Parents completed the CBCL at the baseline assessment as well at 6 months, 9 months, 12 months, 18 months, and 24 months post-visit. The CBCL consists of 145 statements about child behavior over the past six months, rated on a three-point scale: never, sometimes true, mostly true. CBCL scores were computed for maternal and paternal report separately. Scores were available in the following proportions across time points: 78% of mothers and 49% of fathers (baseline), 66% of mothers and 48% of fathers (6 month follow-up), 65% of mothers and 48% of fathers (9 month follow-up), 55% of mothers and 33% of fathers (12 month follow-up), 38% of mothers and 26% of fathers (18 month follow-up), and 28% of mothers and 22% of fathers (24 month follow-up).

We utilized the higher-order Internalizing and Externalizing scales and the lower-order scales measuring Aggression and Rule-breaking. We also utilized the following DSM-oriented subscales, which were derived using the items most closely linked to DSM-IV disorder criteria: Conduct disorder and Oppositional Defiant Disorder.

We also computed and utilized the callous-unemotional subscale created by Willoughby and colleagues (2011, 2014). As Willoughby used the preschool version of the CBCL and we utilized the school-age version, we are able to recover four of the five items in the original scale ("Does not seem to feel guilty after misbehaving," "Seems unresponsive to affection," "Shows little affection toward people," "Shows too little fear of getting hurt"). The item "Punishment does not change behavior" is not included in the school-age CBCL and was excluded. Our scale demonstrated somewhat lower Cronbach's alphas than the original five-item scale (Motherreport: baseline $\alpha = .27$, 6 months $\alpha = .46$; Father-report: baseline $\alpha = .39$, 6 months $\alpha = .35$) (Willoughby et al., 2011).

Children's Behavior Questionnaire (CBQ). Participating mothers and fathers completed the Children's Behavior Questionnaire (CBQ; Rothbart, Ahadi, Hershey, & Fisher, 2001), which was designed to measure temperament in children aged 3 to 7 years. Parents completed the CBQ at the baseline assessment as well at 6 months, 9 months, 12 months, 18 months, and 24 months post-visit. The CBQ consists of 195 items rated on a seven-point scale from "extremely untrue" to "extremely true." The CBQ yields three higher-order subscales (Surgency, Negative Emotionality, and Effortful Control) and lower-order subscales measuring more homogenous facets of temperament. For our analyses, in addition to the higher-order subscales, we chose to examine lower-order scales based on theoretically expected associations with guilt and empathic responding. We utilized the following subscales consistent with empirical and theoretical associations with empathy and guilt: Anger, Attentional focusing, Sootheability, Fear, High intensity pleasure, Impulsivity, Inhibitory control, Sadness, and Shyness. Scores were available in the following proportions across time points: 79% of mothers and 50% of fathers (baseline), 66% of mothers and 48% of fathers (6 month follow-up), 65% of mothers and 48% of fathers (9 month follow-up), 55% of mothers and 33% of fathers (12 month follow-up), 38% of mothers and 26% of fathers (18 month follow-up), and 28% of mothers and 22% of fathers (24 month follow-up).

Experimenter ratings of child traits. At the conclusion of the initial laboratory visit, experimenters rated the child's global behavior across the visit. Behaviors were rated on a five-point Likert scale. We utilized the three higher order dimensions: Effortful Control, Negative Emotionality, and Positive Emotionality (see Vroman, Lo, & Durbin, 2014). Post-visit ratings were available for 99.6% of children.

Alabama Parenting Questionnaire (APQ). Mothers and fathers completed the Alabama Parenting Questionnaire (APQ; Frick, 1991) at the baseline assessment. The APQ is a 42-item questionnaire assessing parenting practices across five dimensions associated with risk for conduct problems: Poor Monitoring/Supervision, Corporal Punishment, Inconsistent Punishment, Positive Parenting, and Parental Involvement. Scores were available for 78% of mothers and 49% of fathers.

Data Analytic Plan

Aim 1: Describe variations empathic responding in a morally relevant scenario. We utilized two analytic strategies to describe variations in empathic responding. All participants with Picture Tearing data were retained for factor and cluster analysis, regardless of whether questionnaire outcome data was present. However, only those participants with CBCL, CBQ, or APQ data were included in follow-up external validation analyses.

Factor analysis. First, we utilized a variable centered approach and conducted an Exploratory Factor Analysis in Mplus version 7.2 (Muthen & Muthen, 1998-2012) utilizing a principal axis approach with oblimin rotation to account for the fact that the underlying dimensions are not orthogonal to one another. Picture Tearing task variables from all three epochs were utilized in the EFA. Those that demonstrated no associations with other task variables were excluded from analyses. Eigenvalues, scree plot, and parallel analysis were all utilized to estimate the most likely number of factors. We retained the three factor structures best supported by analyses and tested the fit of these models using Confirmatory Factor Analysis in Mplus (Muthen & Muthen, 1998-2012). We also estimated a three-factor model in which factors and loadings were chosen based on theoretical assumptions about the underlying structure of task behavior (see Table 4; Hypothesis 1.1). Fit indices for each model were compared. We

created composite variables for further analysis based on the factor structure of the best-fitting model.

Cluster analysis. Second, we utilized a person centered approach by performing a cluster analysis on the Picture Tearing data. Cluster analysis included all of the coding variables from all three epochs. To derive clusters, we used a two-step method recommended by several researchers (e.g. Hair et al., 1994; Milligan, 1980; Steinley, 2003; Swogger, 2007). We initially derived clusters using Ward's hierarchical agglomerative method of cluster analysis in SPSS (IBM version 22.0). To determine the appropriate number of clusters, we graphed the agglomeration coefficients against the stage number and used the "elbow method" to determine the stage at which agglomeration coefficients ceased to increase substantially from stage to stage (e.g. Swogger, 2007).

We also computed the gap statistic, which indicates whether the pooled squared distances of points within an increasing number of clusters are sufficiently different than would be expected based on a random, uniform reference distribution (see Tibshirani et al., 2001). We computed this statistic using the clusterGenomics package (Nilsen & Lingjaerde, 2013) in R (R Core Team, 2013). Based on these analyses, we chose a number of clusters and applied this solution to the data using the k-means cluster procedure (Tan, Steinbach, & Kumar, 2006) in SPSS, with k set as the number of clusters derived from the hierarchical method. We validated the resultant clusters using several methods: (1) collapsing of similar clusters that have fewer than 5% of the sample included to avoid identification of small groups unlikely to replicate in other samples and (2) validation of clusters by comparing them on variables not included in the cluster analysis (i.e., parent-reported temperament and behavior problem, child sex and age, etc.).

This analysis allowed us to examine characteristics that may differ between children who present with certain patterns of behaviors. Overall, we expected that most children will comply with the experimenter's request (Hypothesis 1.2), but that children's compliance may have very different accompanying behaviors (e.g., tearing with glee versus reluctantly and with concern for the second experimenter). We hypothesized that failure to comply could be positively associated with assertiveness (Hypothesis 1.3) and/or extreme levels of expressed empathy (Hypothesis 1.4) in some children. However, failure to comply may also correlate positively with defiance or oppositionality (Hypothesis 1.5), as measured by the child's noncompliant and defiant behaviors. Compliance in tearing the picture in the absence of empathic statements could also demonstrate a positive association with fearlessness (Hypothesis 1.6), as these children are unlikely to feel concerned about the experimenter's feelings, and may even enjoy tearing the picture.

Aim 2: Examine associations between observed moral behavior and child characteristics theoretically linked to moral development and empathy. Utilizing composites derived from factor analysis and a variable indicating cluster membership, we examined the pattern of bivariate correlations between factors, cluster membership, and parentreported problem behaviors. We excluded variables that did not have a significant zero-order association with the dependent variable of interest. The general strategy for regressions is described below; in all cases, variables described in each step were only included in the regression if they had a significant bivariate correlation with the dependent variable of interest. In all analyses, mother-reported variables and father-reported variables were included and analyzed separately.

Hierarchical regressions were run separately predicting total externalizing problems, aggression, CU traits, Rule-breaking, ODD symptoms, CD symptoms, and total internalizing problems. In all analyses, we first entered child sex and age. Then, we entered Picture Tearing variables (either behavioral composites or dummy coded contrast variables describing cluster membership) into step two. The purpose of these initial regressions was to determine if behaviors during Picture Tearing predict externalizing and internalizing problems over and above demographic variables that are themselves associated with variations in problem behaviors.

The next set of regressions more stringently evaluated whether Picture Tearing behaviors predicted problem behaviors over and above other child and environmental characteristics that are theoretically and empirically linked to externalizing pathology (see Figure 3). We included additional predictors only if the independent variables accounted for significant variance in the outcome in the first set of regressions. In these additional analyses, we first entered child sex and age; subsequent steps in the regression will include emotion recognition and verbal ability (step 2), child temperament traits (step 3), and parenting dimensions (step 4). Finally, Picture Tearing behaviors will be entered (step 5). The order of the sets of variables was determined by reference to theoretical ideas regarding their casual primacy for the dependent variables. We chose to enter emotion recognition and verbal ability next as these factors may fundamentally impact the degree to which the child understands the moral scenario of the task. Finally, we entered child traits, and then parenting factors; we also reversed the order of steps 3 and 4 in subsequent analyses, as it could be argued that parenting dimensions may be more casually prior to our dependent variables of interest than child traits. Picture Tearing variables were entered

last to determine whether these task-specific behaviors are incrementally predictive beyond other characteristics of the child and his or her environment.

Importantly, the Picture Tearing task was collected as part of battery of 16 tasks, which were designed to measure other traits of interest in this study, including fear proneness and effortful control, as well as compliance. Given that the other 15 tasks do not involve moral dilemmas, coding data from these tasks were used to derive a composite score reflective of the child's typical levels of compliance and fear proneness. We included these lab-based trait composites in our hierarchical regression analyses (assuming they had a significant zero-order correlation with the dependent variable of interest) to determine whether a child's pattern of behavior during Picture Tearing could be accounted for wholly by that child's typical levels of compliance (or fear proneness), or whether compliance (or anxious arousal) in Picture Tearing is incrementally useful in predicting other behaviors (e.g., externalizing). We expected that compliance and fear proneness observed specifically during Picture Tearing would incrementally predict externalizing behaviors, above and beyond typical levels of compliance and fear proneness. Specifically, we expected that noncompliance would be positively associated with externalizing behaviors (Hypothesis 2.1) and fear-proneness, as indexed by Anxious Arousal during the task, would be negatively associated with externalizing behaviors (Hypothesis 2.2).

Irrespective of compliance with the request to tear the picture, we expected that empathy would be strongly positively associated with appropriate moral socialization, operationalized as an absence of significant externalizing problems (Hypothesis 2.3). This is consistent with Kohlberg's (Kohlberg & Kramer, 1969) theories of moral development, which suggest that strict obedience to authority figures is normative for younger children. Consequently, we expected that high anxious arousal and empathy in the Picture Tearing task would be suggestive of

normative moral development and thus correlate with few externalizing problems, while low empathy and low arousal would be suggestive of aberrant moral socialization and accordingly, predict high levels of externalizing problems (Hypothesis 2.2, 2.3).

A summary of our hypothesized associations is provided in Table 3. We discuss these hypotheses in more detail below.

Demographic Variables. In our study, age served as a proxy for developmental period. Across research, older children (5 to 6 years old) typically demonstrate more compliance with the experimenter commands across a variety of tasks than do younger children (3 to 4 years old), including fewer instances of direct defiance (Braungart-Rieker, Garwood, & Stifter, 1997; Kochanska & Aksan, 1995; Kochanska, Aksan, & Koenig, 1995; Laupa, 1991; Laupa & Turiel, 1991). Given the demand characteristics of the situation and previous obedience research (e.g. Shanab & Yahya, 1977), we expected that most older children will comply with the experimenter's request. However, older children are also more likely to demonstrate more sophisticated perspective taking and higher levels of empathic internalization (Borke, 1971, 1973; Kochanska, 1993), and are more likely to view commands to harm others as illegitimate (Damon, 1977; Laupa, 1991, 1994). Thus, we expected that irrespective of their choice to comply, older children would demonstrate higher levels of empathy and anxious arousal, including more empathic statements, than younger children (see Table 3; Hypothesis 2.5, 2.6). Given the compliance research as a whole, we also expected that, given they display more empathy than younger children, older children would be modestly less likely to comply with the experimenter's request than younger children (Hypothesis 2.4).

The potential for sex differences was more exploratory, as the literature does not support a clear position on whether gender significantly predicts compliance or empathy in children.

While many studies report higher levels of compliance for girls (e.g. Kochanska, 2002; Kochanska, Coy, & Murray, 2001; Kochanska, Tjebkes, & Forman, 1998), other research has found no gender differences (e.g. Abe & Izzard, 1999; Braungart-Rieker, Garwood, & Stifter, 1997; Higbee, 2012). Thus, we tentatively expected that girls would exhibit greater compliance than boys, but that this difference would be small in magnitude (Hypothesis 2.7). Similarly, while some studies report that girls demonstrate higher levels of empathy and internalization (e.g. Garaigordobil, 2009; Hoffman, 1977; Kochanska, DeVet, Goldman, Murray, & Putnam, 2008), other research has found that gender differences are either nonsignificant or an artifact of assessment method (e.g. Eisenberg & Lennon, 1983).

A review by Rose and Rudolph (2006) supports the contention that sex differences may become more apparent across development, such that girls demonstrate greater empathy than boys in older school-age or adolescent samples, but effect sizes are small or nonexistent for younger school-age samples. Older children, particularly those of pubertal or post-pubertal age, may experience socializing pressures differentially by sex and may be basing their moral decisions on different factors, whereas younger children may reason more uniformly. This notion is supported by evidence in adolescents and adults that women are more likely to identify "care" as a moral issue (i.e. compassion for the other, providing for others' needs), while men are more likely to identify "justice" as the central dilemma (i.e. fairness, reciprocal rights) (Gilligan & Attanucci, 1988). Similarly, in their sample of 6 to 18-year-olds, Shanab & Yahya (1977) asked children why they chose to comply with the experimenter's request to shock a peer; girls were more likely to state that they were obeying orders, while boys were more likely to express that punishment is beneficial for learning. Collectively, these data suggest that, by middle childhood, boys and girls appear to preferentially focus on different aspects of moral dilemmas

and employ different reasoning. Given the available evidence, we expected to find sex differences in empathy and anxious arousal, of small magnitude (Hypothesis 2.8, 2.9).

Cognitive characteristics. We also examined the role of cognitive maturity in predicting compliance, arousal, and empathy. While some effect of cognitive maturation will be captured by age, even children of the same age display individual variation in cognitive capabilities. We expected that adequately developed cognitive skills would be most strongly associated with perspective taking, given that this empathic skill is predicated on a child's ability to recognize and represent the emotions of others and predict the emotions that might result from a given situation (Pons, Harris, & de Rosnay, 2004; see Table 3). Moreover, poor emotion recognition, especially for fear and sadness, is robustly linked to CU traits, low empathy, low anxious arousal, and aggression (e.g. Blair & Coles, 2000; Blair, Colledge, Murray, Mitchell, 2001; Dadds, Perry, Hawes, et al., 2006; Munoz, 2009; Blair, Budhani, Colledge, Scott, 2005; Fairchild, van Goozen, Calder, et al., 2009). Thus, we expected that children who performed poorly on the baseline trial of our emotion recognition task would also demonstrate low levels of anxious arousal and empathy during the Picture Tearing task (Hypothesis 2.10, 2.11). Given the relationship between general intelligence and the ability to mentally represent complex ideas, we also investigated the potential correlation between perspective taking and general intelligence (as assessed by the PPVT; Hypothesis 2.12, 2.13).

Temperamental characteristics. We expected that a number of child characteristics would be associated with compliance, empathy, and arousal during the Picture Tearing task (see Table 3). Broadly, we expected that compliance, empathy, and anxious arousal (fear-proneness) in Picture Tearing would be positively related to characteristics associated with adequate socialization (as operationalized by lower levels of externalizing problems). There is extensive

support linking fearlessness, low anxious arousal (e.g. Kochanska, 1991, 1993, 1995;

Kochanska, Coy, & Murray, 2001), deficient empathy, and high levels of externalizing problems (e.g. Frick et al., 2014; Frick & White, 2008; Frick, Ray, Thornton, & Kahn, 2014; Frick et al., 1999; Kochanska, Coy, & Murray, 2001) to poor moral internalization and poor compliance (or oppositionality). Specifically, fearlessness has been linked to low arousal during a transgression and low or absent guilt after the transgression (Kochanska, 1995; Kochanska, Gross, Lin, & Nichols, 2002; Rothbart, Ahadi, & Hershey, 1994).

With respect to compliance with the request to tear the picture, we predicted that high levels of externalizing problems (especially ODD symptoms and noncompliance) would be positively associated with noncompliance with the experimenter's request (Hypothesis 2.1). We also expected that noncompliance would be positively associated with tempermental characteristics associated with externalizing behaviors (Hypothesis 2.18), especially fearlessness (Hypothesis 2.14), but negatively associated with sadness (Hypothesis 2.21).

With respect to moral emotions, we expected that observed anxious arousal and empathy would be inversely associated with externalizing problems (e.g. CU traits, aggression, rule-breaking, ODD symptoms; Hypothesis 2.2, 2.3) and temperamental characteristics theoretically and empirically associated with externalizing (e.g. anger, taking pleasure in highly stimulating activities; Hypothesis 2.19, 2.20), especially fearlessness (Hypothesis 2.15, 2.16). Conversely, we expected that anxious arousal and empathy/guilt would be positively associated with sadness (Hypothesis 2.22, 2.23). Taken as a whole, these findings would be consistent with a preponderance of findings associating moderate fearfulness and anxious arousal with appropriate moral internalization and guilt (e.g. Kochanska, 1995; Kochanska, Coy, & Murray, 2001; Kochanska, Gross, Lin, & Nichols, 2002; Stifter, Cipriano, Conway, & Kelleher, 2009).

However, extreme levels of fearfulness have also been linked to low empathy, given that these children may be oriented more towards personal distress than the distress of others (Eisenberg et al., 2010). Thus, we anticipated a curvilinear relationship between empathy and fearfulness, such that children with moderate levels of fearfulness demonstrate the most empathy, while children at either extreme demonstrate low levels of empathy (Hypothesis 2.17). If fearlessness was related at a zero-order level to the outcome of interest, this was tested by entering a nonlinear term for fearfulness after a linear term in a hierarchical multiple regression. Furthermore, we expected that high fearfulness would correlate positively with statements indicative of personal distress during the Picture Tearing task.

The expected pattern of associations for facets of effortful control was partially exploratory. In general, research supports a positive relationship between compliance with orders and effortful control and a negative relationship with impulsivity (Kochanska, Coy, & Murray, 2001; Kochanska et al., 1997). This relationship is especially strong for "Do" tasks, in which the child is being asked to perform an action. However, these findings are drawn almost exclusively from benign orders (e.g. clean up) rather than orders to do harm (e.g. tear the picture), and the tasks involved are typically sustained behaviors (e.g. clean up the room, carry marbles one by one) rather than single actions (e.g. tear the picture). Results from the obedience literature suggest that the pull to obey an authority figure will, for most children, outweigh the pull to behave morally (e.g. Shanab & Yahya, 1977). Thus, we expected that effortful control and its facet of inhibitory control would be positively related to compliance, while impulsivity would inversely predict compliance (Hypothesis 2.27).

Additionally, effortful control capacities have been positively linked to high moral internalization and a developed conscience and negatively linked to disobedience (e.g.

Kochanska & Knaack, 2003; Kochanska, Murray, & Coy, 1997), which suggests that children in our study with high effortful control may also demonstrate higher levels of moral internalization, which would suggest a reduced likelihood of committing an immoral act. Given the available literature, we predicted that effortful control would also be positively associated with anxious arousal, empathy, and guilt (Hypothesis 2.28, 2.29).

Parenting factors. Finally, we investigated the role of broader contextual factors in predicting the child's ability to comply with requests and overall moral internalization. Specifically, we examined effects of the parent-child relationship, given that aspects of this relationship have demonstrated numerous empirical associations with compliance and conscience development. On the whole, the literature suggests that high levels of parental control, particularly power assertion, harsh discipline, and corporal punishment, are inversely related to both compliance (Abe & Izard, 1999; Braungart-Rieker, Garwood, & Stifter, 1997; Crockenberg & Litman, 1990) and conscience (empathy, guilt) (Kochanska, 1991, 1993, 2002). A warm parent-child relationship can even suppress the deleterious effects of fearlessness on internalization (Kochanska, 1995), while power assertion demonstrates few associations with conscience for highly fearless children (Kochanska, 1991). Thus, in general, we expected to find significant negative associations between observed compliance, anxious arousal, and empathy and parent-reported poor monitoring, corporal punishment, and inconsistent punishment (see Table 3; Hypothesis 2.30, 2.31, 2.32). Conversely, we expected that compliance, empathy, and guilt will be positively associated with parent-reported positive parenting and involvement (Hypothesis 2.33, 2.34, 2.35). However, we also hypothesized that these relationships would be moderated by child fearlessness, such that the relationship between power assertion and child empathy would be significant only for moderately or highly fearful children

Aim 3: Examine predictive associations between compliance, arousal, and empathy and later externalizing behaviors. Finally, we examined whether observationally assessed compliance, arousal, and empathy at baseline would predict outcomes at 6 months, 9 months, 12 months, 18 months, and 24 months after the initial assessment. We were especially interested in the prediction of externalizing pathology, given its robust empirical associations with low empathy, poorly developed conscience, and disorders of empathy (i.e., CU traits) (e.g. Asendorpf & Nunner-Winkler, 1992; de Wied, van Boxtel, Matthys, and Meeus, 2011; Frick et al., 2003; Frick & Morris, 2004; Frick, Ray, Thornton, & Kahn, 2014; Frick & White, 2008; Jolliffe & Farrington, 2004). Given the accumulation of evidence, we predicted that empathy and arousal during the Picture Tearing task will predict externalizing problems (e.g. CU traits, aggression, rule-breaking, ODD symptoms, CD symptoms) at all follow-up time points, over and above the variance associated with baseline levels of these behaviors. To assess discriminant validity, we also examined the relationship between empathy and arousal and later internalizing problems. In all analyses, mother-reported variables and father-reported variables were included and analyzed separately.

To investigate these questions, we analyzed the patterns of change in these variables over time using growth curve analysis estimated in HLM (Scientific Software International Inc.). We fit a two-factor linear growth model using multi-level modeling in which the predictive variables of interest from Picture Tearing (dimensions derived from EFA and groups from k-means clustering) were the primary independent variables of interest, and the dependent variables were repeated measures of EXT in children across the multiple follow-up waves. Level-1 of the multi-level model included the within-participant change in EXT over time, and we modeled individual differences in the intercept of EXT (baseline values) and in the linear age slope of

EXT. At level-2 of the multi-level model entered between subjects predictor variables (i.e., behavioral dimensions and clusters from Picture Tearing; other covariates of interest identified from hierarchical regression analyses described above, such as child age and sex). In this way, we assessed whether baseline observed guilt, empathy, and compliance reliably relate to individual differences in change in externalizing behaviors over a period of approximately 2 years. Depending on model fit, we also investigated the possibility that the data is best represented by a nonlinear equation (i.e., quadratic or cubic change in externalizing problems). Specifically, we expected that noncompliance would be positively related to later externalizing behaviors (Hypothesis 3.1), while empathy and anxious arousal would be negatively related to later externalizing behaviors (Hypothesis 3.2, 3.3).

RESULTS

Recoding of Picture Tearing Variables

Across variables, there was a relatively low endorsement of behaviors, which resulted in skewed distributions for all of the coding variables and wide ranges of responses. Absolute values for skewness and kurtosis ranged from 0.29 to 32.23. Only seven variables had absolute values for both skewness and kurtosis falling below 2.00, which is generally recognized as an acceptable cutoff, although there is disagreement in the literature as to a standard rule of thumb (e.g. Jones, 1969; see Table 6). To decrease variable skew, we recoded each variable by coding ranges of responses into single codes. We utilized histograms and percentiles to ensure cut points that would provide relatively even groupings with equal percentages of total participants in each grouping, where equal groupings were possible while still maintaining adequate variability in the codes. However, very low base rates of some behaviors prevented us from creating equal groups for some variables; in some cases, over half of the participants scored a 0 for a given variable, so equal groupings would have been impossible without collapsing unlike codes. In other cases, equal groupings would have reduced variability in the final code by obscuring differences between participants who scored at the upper range of the distribution. Cut points differed by variable, given the incredible variation in distributions across the coding variables. Thus, applying a standard recoding solution to each variable was unreasonable. Rather, cut points and groupings were chosen to maintain variability while reducing skewness and kurtosis. For example, Empathy was recoded such that responses greater than 0 but less than or equal to 1 were coded as 1, responses greater than 1 but less than or equal to 3 were coded as 2, and responses greater than 3 were coded as 3. In this case, over half of participants scored a 0 for Empathy, so we were unable to create equal cut points without erasing the

differences between any number of statements above 0. Instead, we examined all responses from 0 to the maximum and chose relatively equal groupings *within this range*, so that the frequencies of the codes above 0 were relatively equal. In this way, we were able to reduced skewness and kurtosis while preserving potentially important distinctions between children who make only 1 empathic statement and children who make 3 or more. After recoding, almost all of the variables had absolute values for skewness and kurtosis below 2.00, with values for only two variables (Blaming Self and Apologizing) exceeding 3.00 (see Table 7).

Responses coded during Epoch 2 demonstrated very large skewness (with no more than 30 participants exhibiting a codeable response for any Epoch 2 variable). In addition, there were not meaningful conceptual differences between certain behaviors (e.g. empathy) exhibited in different epochs. Specifically, there were no theoretical reasons to assume that making empathic statements during Epoch 1 would have different correlates than making empathic statements during Epochs 2 or 3. Thus, variables coded across all three epochs were averaged into composites for further analyses (Empathy, Hesitation, Defiance, Social Reference, Laughing, Smiling). It is notable that these composites correlated between .40 and .98 with the variables that comprised them.

Recoded and composite variables were retained for all future analyses. The variables Squirming, Lip Biting, Crossing Arms, Hiding, and Hunched Shoulders were dropped from factor analyses due to very low rate of occurrence across epochs, even after recoding, few to no significant correlations with other behaviors, and/or low reliability between coders.

Bivariate Correlations Among Recoded Variables

Correlations were examined between the recoded variables (see Table 8). The magnitude of most correlations ranged from small to moderate. The associations between variables largely

conformed to expectations, with some exceptions. Overall, verbalizations associated with Empathy, Hesitation, and Defiance were associated with lower task compliance, as we had anticipated (Hypothesis 2.36), whether noncompliance was measured by how many prompts the experimenter gave before the child complied (Number of Prompts) or by the time elapsed between the first prompt and the child's compliance or experimenter's giving up (Latency to Tear). Only Empathy (r = -0.22) and Defiance (r = -0.60) predicted less absolute compliance (whether the child tore the picture or not) and less complete compliance (tearing the picture a small amount as opposed to a large amount) (r = -0.27 for Empathy, r = -0.60 for Defiance). Compliance itself (choosing to tear the picture) was also positively associated with Social Referencing (r = 0.19), expressions of positive affect (r = 0.18 for Laughing, r = 0.28 for Smiling), and Enjoyment in the task (r = 0.16). Only empathic statements and nonverbal signs of discomfort (Hunched Shoulders, Squirming, Gaze Avoidance) were associated with higher overall guilt (respectively r = 0.17, r = 0.16, r = 0.24) (Hypothesis 2.36).

We did not have specific expectations for which variables would predict responses to the victim, as this part of the investigation was largely exploratory. Interestingly, Silence and Lying were both associated with more Tearing (r = 0.27 for Silence, r = 0.28 for Lying) and fewer Prompts (r = -0.34 for Silence, r = -0.22 for Lying), while more Tearing was independently correlated with Blaming the Other (r = 0.24), Blaming the Self (r = 0.13), and Gaze Avoidance (r = 0.17). Overall guilt was negatively associated with Blaming the Self (r = -0.16), suggesting that children who expressed more guilt were also less likely to take full responsibility for tearing the picture ("I did it" as opposed to "She told me to do it"). It is possible that children who

more motivated to alleviate the aversive internal state (guilt) by displacing some of the blame from themselves.

Notably, this pattern of correlations does not meaningfully or substantially differ from the pattern of correlations observed between the variables when separated by epoch.

Overall, children who expressed empathy and defiance were less likely to comply with the task at all, while expressions of empathy, defiance, and hesitation were all related to slower, less enthusiastic, less thorough compliance. Only expressions of empathy were associated with higher guilt when confronted by the victim. Furthermore, more thorough compliance (more tearing) was associated with silence, lying, blaming the other or self, and gaze avoidance. Guilt was actually associated with less blaming the self. Overall, these data support that meaningful coherence can be found within task behaviors, and that this coherence is in accordance with theorized associations (e.g. empathy predicting more guilt). This suggests that children are responding to the Picture Tearing task as though it were a real moral dilemma. Thus, the Picture Tearing task appears to be a useful naturalistic scenario in eliciting moral emotions and morallyrelevant behavior from young children.

Associations Between Task Behaviors and Demographic Variables and Environment Characteristics

While above we had examined the associations among single task variables, we were also interested in the associations between task variables and other child and environmental characteristics that have been empirically and theoretically linked to compliance, moral decision making, and moral emotions. We first examined bivariate correlations between task variables (separated by epoch, totals, and factors) as well as demographics, temperament traits, and other variables of interest (see Table 9, Table 10).

Age and sex. As expected, age was significantly and positively associated with empathic statements (epoch 1) (r = 0.14), which is consistent with previous literature suggesting that older children demonstrate better perspective taking than younger children (Hypothesis 2.6). Older children were more likely to express hesitation and to exhibit social referencing behavior, but age was unrelated to defiant statements. Unexpectedly, age was not correlated with whether the child tore the photo (absolute compliance) or with latency to tear (Hypothesis 2.4). Instead, age was negatively associated with the number of prompts given (r = -0.12) and positively associated with how much the photo was torn (r = 0.15), contrary to our hypothesis that older children would, by virtue of expressing more empathy, also comply less (Hypothesis 2.4). Thus, in our sample, older children were not necessarily more likely to comply at all, but they did comply more quickly and more extensively than younger children. Older children also demonstrated more discomfort when facing the victim, but did not appear guiltier overall, which only partially supports our expectation that older children would display more anxious arousal (Hypothesis 2.5).

To further examine the effect of age, we created a dichotomous variable with children coded as younger (3 to 4 years) or older (5 to 7 years) and conducted independent samples *t* tests to examine differences between the groups. This division was chosen because it has been used as a proxy for developmental stage in previous moral development research (e.g. Braungart-Rieker, Garwood, & Stifter, 1997; Kochanska & Aksan, 1995; Kochanska, Aksan, & Koenig, 1995; Laupa, 1991; Laupa & Turiel, 1991). Consistent with correlational results and our hypotheses, older children had higher means for empathy (epoch 1; *p* = .03), hesitation (epoch 1; *p* = .04), total social referencing (*p* = .03), and relief for the copy (*p* < .05) (Hypothesis 2.6). Older children were also surprisingly higher in overall positive affect (*p* < .001) and enjoyment

in tearing (p < .01), and were more likely to blame the other person (p < .01) but less likely to take the blame themselves (p = .02).

While we hypothesized that girls might show slightly more empathy and anxious arousal and less noncompliance than boys (Hypotheses 2.7, 2.8, 2.9), findings for sex were largely consistent with the body of research suggesting that sex differences are largely nonexistent at this age range. Sex was significantly correlated only with Squirming (epoch 3) and Relief for the copy. Independent sample *t* tests similarly found that girls demonstrated more squirming (epoch 3; p < .01) and Relief (p = .01) than boys, with no other significant differences in group means.

Cognitive characteristics. We also examined the relationships between task variables and cognitive characteristics, specifically verbal intelligence and performance on an emotion recognition task. Consistent with our expectations, verbal intelligence (as measured by total PPVT score) was modestly, positively associated with total empathic statements (Hypothesis 2.13) (r = 0.16). It was not correlated with anxious arousal, however (Hypothesis 2.12).

Furthermore, we found a number of associations between task variables and emotion recognition scores that were consistent with theoretical links between poor emotion recognition ability and poor perspective taking. As we had expected, total empathy was modestly positively related to total correct emotions identified (how many emotions the child correctly identified out of the six emotions presented) (r = 0.13), and was also independently, negatively associated with incorrectly identifying sad faces (Hypothesis 2.11) (r = -0.14). Furthermore, we observed the same pattern of associations for Social Referencing behavior, one marker of anxious arousal (Hypothesis 2.10). However, emotion recognition scores were not associated with overall guilt.

Parenting characteristics. We also examined the association between task variables and parent-reported characteristics of their own parenting. Consistent with our expectations, we

did observe some significant associations with negative parenting characteristics and child defiance and noncompliance and between (positive parenting and adequate socialization behaviors. Specifically, mother-reported involvement was associated with more child expressions of empathy in the first epoch (Hypothesis 2.35) (r = 0.14). Mother-reported poor monitoring was associated with more child defiant statements (r = 0.14) and more noncompliance (r = 0.19) (Hypothesis 2.30), but was also unexpectedly associated with more social referencing (r = 0.15) (Hypothesis 2.31). Similarly, and also unexpectedly, we observed unexpected positive associations between father-reported poor monitoring (r = 0.18) and corporal punishment (r = 0.18) and child empathic statements (Hypothesis 2.32). Results concerning monitoring should be interpreted cautiously, as most of the items (e.g. "Your child stays out in the evening past the time he/she is supposed to be home") are more applicable to older children than to children the age of those in our sample (3 to 7 years). Though this scale is less appropriate for younger children, we included it because poor monitoring is robustly associated with externalizing behaviors in other research (e.g. Beyers et al., 2003), and because it does include items potentially applicable to children the age of our sample, particularly at the older end of the range (e.g. "You leave the house without telling your child where you are going"). Furthermore, the scale was validated in other research using samples of children as young as 6 years of age, and thus may have utility for the oldest children in our sample (Shelton, Frick, & Wooton, 1996).

Overall. Overall, older children expressed more empathy and hesitation than younger children (Hypothesis 2.5, 2.6), but despite this, were actually slightly more likely to comply with the experimenter's request (Hypothesis 2.4). Sex effects were limited, with girls demonstrating slightly more discomfort in the form of squirming than boys, and seeming slightly more relieved
when the copy of the destroyed picture was revealed (Hypothesis 2.8). Sex differences were not observed in empathy (Hypothesis 2.9) or noncompliance (Hypothesis 2.7).

Verbal intelligence and emotion recognition ability were both associated with more empathic statements (Hypothesis 2.11). Accurate emotion recognition was also associated with more anxious arousal in the form of social referencing, when the child looks at the experimenter presumably for reassurance, but was not associated with other markers of anxious arousal (Hypothesis 2.10).

Positive parenting for mothers was associated with more child empathic statements (Hypothesis 2.35). Poor monitoring for mothers was related to more defiance and noncompliance on the part of the child (Hypothesis 2.30), but also unexpectedly more empathic statements (Hypothesis 2.32). Similarly, poor monitoring for fathers was unexpectedly related to more guilt (including both global guilt and empathic statements; Hypothesis 2.32) on the part of the child.

Overall, these data support that task behaviors derived from the coding scheme are reflective of meaningful differences in child standing on other internal and external characteristics. This suggests that the coding scheme has utility in measuring behaviors that are of theoretical interest in themselves, and also relate to a child's behavior outside of the artificial laboratory setting.

Exploratory Factor Analysis

We conducted principal axis factor analyses utilizing oblimin rotation in MPlus version 7 (Muther & Muther, 1998-2011). While seven factors had eigenvalues above 1.0, evaluation of the initial scree plot suggested that no more than four factors would be appropriate. We then conducted a principal axis exploratory factor analysis in MPlus constraining the solutions to one,

two, three, four, and five factors. Chi square difference analyses indicated that fit differed significantly between all five solutions (p < .001) and that the five factor model demonstrated the best fit, although fit statistics were modestly below typically used cutoffs (Hooper, Coughlan, & Mullen, 2008). We also performed a parallel analysis, which compared a scree plot generated from the actual data to what would be expected if the data were randomly generated. The parallel analysis suggested that only four factors were larger than could be expected by chance, which recommended a four factor structure. Empirically derived factor structures for the solutions indicated by the scree plot and parallel analysis are provided in Table 11. We also retained the three factor solution, given that we had hypothesized a three factor solution. This allowed us to compare our hypothesized number of factors with the factors suggested empirically.

Confirmatory Factor Analysis of Empirically Derived Factors

To further examine the differences between the potential factor solutions, we conducted confirmatory factor analyses using ML estimation. We compared fit between the three, four, and five factor solutions (loadings are described in Table 13). Indicators that did not load above .30 on any factor (Social Referencing, Blaming Self, Apologizing, and Lying for the three, four, and five factor models and Gaze Avoidance for the three factor model) were dropped from analyses. Notably, these variables also demonstrated few or very modest correlations with other indicators (see Table 8).

Investigation of the fit indices (see Table 12) suggested that fit differed substantially between factor solutions, with the five factor solution demonstrating the best fit overall according to standard fit indices. Notably, the fit statistics for all three were outside the bounds of typically used cut-off scores. Given that fit indices are affected by sample size and non-

normality, and that our sample size was relatively small and our variables somewhat skewed, the standard cut-offs may not be optimal in evaluating this data (Nye & Drasgow, 2011). Fit for the three factor model was not substantially improved by modifications suggested by modification indices. The four and five factor models both demonstrated improved fit with modifications to the latent structure. This was contrary to our expectation of a three factor model conforming to Noncompliance, Anxious Arousal, and Empathy (Hypothesis 1.1).

In comparing the modified four and five factor structures, the four factor structure demonstrated the best fit, though change in fit indices was modest between the four and five factor models. The four factor model is more parsimonious. In addition, parallel analysis (discussed above) suggests that only the first four factors are sufficiently larger than would be expected due to chance. However, the five factor structure is consistent with theoretical associations. The four-factor model places both empathy-related and arousal-related variables into a single factor, though theoretically they involve related but distinct aspects of empathic responding. Specifically, our operationalization of empathy involves statements made by the child, which indexes perspective taking (cognitive empathy), while arousal-related variables should be more closely related to affective empathy (Kochanska, 1993). These two facets of empathy are overlapping but not identical. Given that the five factor model demonstrates the best fit theoretically, and that the increment in fit observed in the four-factor model is very modest, we chose to retain the five-factor model for further analyses. This model is described in greater detail below.

Confirmatory Factor Analysis of Rationally Derived Factors

We also fit a model in accordance with our hypothesized (rational) factor structure (see Table 11) with three latent factors representing Compliance, Arousal, and Empathy. The initial

model did not converge. The model failed to converge even when items with low factor loadings were dropped. In comparing the structure of this rationally derived solution to the empirically derived solutions, it appears that our initial hypotheses for Arousal conflated a tendency to display positive affect (Smiling, Laughing) with a tendency to display discomfort (Gaze Avoidance, Silence, Squirming). Our hypotheses assumed that all expressions of discomfort, whether positively or negatively valenced, were expressions of the same underlying latent variable, anxious arousal (Hypothesis 1.1). However, empirically, expressions of positive affect and expressions of avoidance/discomfort (Gaze Avoidance, Silence, Squirming) corresponded to separate latent factors. While smiling and laughing in this task might still represent internal discomfort rather than internal pleasure (factor analysis is agnostic as to this possibility), these expressions appear to be conceptually different from the other variables measuring emotional arousal.

Final Factor Analysis

Given the data above, we chose to apply only the empirically-derived five factor solution in choosing composites for further analysis. The five factor model retains a Noncompliance, Positive Affect, and Avoidance/Withdrawal factor as observed in the four-factor model, but teases apart the Empathy/Guilt factor into aspects that can be labeled as Appeal to Authority (Hesitation, Blame the Other) and Empathy/Guilt (Empathy, How Guilty, Relief at copy). To create composites, we obtained standardized z-scores for each variable, then created averages in accordance with the five-factor structure. Variables making up each factor are listed in Table 11. Factor loadings for the final five-factor model are presented in Table 13.

Summary of Factor Analysis

Overall, the data did not fit well into the three factor structure of Noncompliance, Empathy, and Anxious Arousal/Guilt that we had originally hypothesized (Hypothesis 1.1). Noncompliance did emerge as a factor. However, Empathy and Anxious Arousal did not separate from each other cleanly. Instead, the two major moral emotions cohered, such that empathic statements and guilt emerged together as a Guilt factor. Expressions attributed to Anxious Arousal actually separated into three distinct factors: a Positive Affect factor, which appears to reflect genuine positive affect rather than nervous smiling; an Avoidance factor, which reflects behaviors designed to withdraw from confrontation; and Appeal to Authority, which reflects behaviors designed to elicit reassurance from the experimenter.

While the data did not factor exactly as we had expected, we did observe covariance in behavior that made sense from a theoretical standpoint. Empathy and guilt loaded onto the same factor; while these are theoretically distinct moral emotions, they are also highly correlated, and it thus makes sense that they might reflect the same underlying factor. Furthermore, anxious arousal behaviors did vary together; they were simply reflective of several distinctive types of anxious arousal, whereas we had expected them to reflect one underlying factor. On the whole, our coding scheme does appear to measure a variety of behaviors that conform to important aspects of moral decision making, but the distinctions with respect to anxious arousal are more nuanced than we had first anticipated.

Associations Between Picture Tearing Task Composites and Other Child and Environment Characteristics

While above we had examined the associations among single task variables and other child and environmental characteristics, we were also interested in examining how these

relationships applied to our factor-analysis derived composite variables. We discuss demographic variables, cognitive characteristics, and parenting variables here; the association between task variables and temperament traits and problem behaviors is described in depth below.

Demographic variables. Consistent with correlational results and our hypotheses, older children had higher means for the Guilt composite (r = 0.13) (Hypothesis 2.5, 2.6).

Cognitive characteristics. Consistent with our expectations, verbal intelligence (as measured by total PPVT score) was modestly, positively associated with the Guilt composite (r = 0.14) (Hypothesis 2.12, 2.13).

Furthermore, emotion recognition scores were not associated with the Guilt composite (which includes both guilt and empathic statements), contrary to hypotheses (Hypotheses 2.10, 2.11). This is also contrary to correlational results for single variables, which suggested that accurate emotion recognition predicted more empathic statements. Given that the Guilt composite includes both empathic statements *and* guilt, it appears that accurate emotion recognition is specifically related to empathic statements, such that the children with the best emotion recognition abilities are most likely to verbalize empathy. However, given that our situation represents an obvious transgression, even children not skilled at emotion recognition may understand that the task is morally wrong and feel guilty.

Parenting variables. Specifically, mother-reported involvement was associated with higher scores on the Guilt composite (r = 0.15) (Hypothesis 2.34, 2.35). Mother-reported appropriate punishment was associated with lower scores on the Avoidance composite (r = -0.19), suggesting that appropriate punishment is related to less withdrawal from the child in the face of their own wrongdoing.

Overall. On the whole, associations for the composite variables were fewer than for the variables analyzed separately. Age and verbal ability were associated with more Guilt (Hypothesis 2.5, 2.6, 2.12, 2.13), while emotion recognition was not (Hypothesis 2.10, 2.11). Maternal involvement was also associated with more Guilt (Hypothesis 2.34, 2.35), while maternal appropriate punishment was associated with less Avoidance (Hypothesis 2.34). Overall, these data support that task behaviors derived from the coding scheme are reflective of meaningful differences in child standing on other internal and external characteristics. Thus, task behaviors appear to reflect children's actual standing on moral emotions and decision making behaviors in relation to their peers, rather than simply reflecting artificial behaviors exclusive to the laboratory environment.

Derivation of Clusters

To derive clusters, we used a two-step method recommended by several researchers (e.g. Hair et al., 1994; Milligan, 1980; Steinley, 2003; Swogger, 2007). The first step involves conducting a hierarchical cluster analysis to determine the most appropriate number of potential clusters. As the hierarchical method is heavily influenced by initial cases assigned to each cluster, the second step involves conducting a *k*-means cluster analysis utilizing the number of clusters and centroids derived from step one.

Hierarchical cluster analysis. We initially derived clusters using Ward's hierarchical agglomerative method of cluster analysis in SPSS (IBM version 22.0) using the recoded picture tearing variables. However, instead of using totals collapsed across epochs, we utilized the variables coded separately for each epoch. We also retained the Squirming, Hunched Shoulders, and Social Reference variables that were excluded from factor analysis.

To determine the appropriate number of clusters, we graphed the agglomeration coefficients against the stage number and used the "elbow method" to determine the stage at which agglomeration coefficients ceased to increase substantially from stage to stage (e.g. Swogger, 2007). The number of this stage is then subtracted from the total number of stages to determine an appropriate number of clusters. This method suggested that a four cluster solution would be appropriate. The change in agglomeration coefficients became modest after stage 5, showing a distinct "elbow" in the plot of coefficients (Figure 3). Subtracted from the total number of stages (269), this suggests 4 clusters.

As a check to determining the number of clusters, we also computed the gap statistic, which indicates whether the pooled squared distances of points within an increasing number of clusters are sufficiently different than would be expected based on a random, uniform reference distribution (see Tibshirani et al., 2001). We computed this statistic using the clusterGenomics package (Nilsen & Lingjaerde, 2013) in R (R Core Team, 2013). Then, according to recommendations from Tibshirani et al. (2001), we graphed the number of clusters against the value of the gap statistic for each cluster (Figure 4). As the graph shows, the most well-separated cluster solution occurs at k = 1, i.e. that the data does not contain more than one "well-separated" cluster. However, the gap statistic rises again at k = 3, suggesting that there are at least three less separated "sub-clusters" within the data.

On the whole, these methods suggest that there might be as many as four distinct clusters within the dataset. However, separating the data into multiple clusters should be interpreted cautiously, as the distance between these cluster centers is not large enough to suggest well-defined, maximally separated groups. Thus, it is possible that there are three to four subclusters

within one single cluster. Further analyses will evaluate the characteristics of these subclusters and whether or not there are meaningful differences between them.

K-means cluster analysis. Next, we fit a four cluster solution to the data using the kmeans cluster procedure in SPSS, specifying k = 4, and a three cluster solution. After further analyses, we retained the three cluster solution, described below. Descriptions of clusters in the four factor solution are provided in Appendix B.

Description of clusters in three cluster solution. Clusters 2 and 4 significantly differed only on Latency to tear (d = 2.98) and Relief in seeing the photo copy (d = 0.67). Notably, they did not differ on the constructs theoretically most important to moral decision making (i.e. Empathy, Defiance, Guilt). In addition, the fourth cluster contained only 6% of the sample. To determine whether collapsing clusters 2 and 4 substantially changed their relationships with the other clusters, we computed descriptives for a three cluster solution (with clusters 2 and 4 collapsed). We then conducted a one-way ANOVA using cluster membership from the new three-cluster solution and conducted post-hoc comparisons of significant ANOVAs using Tukey's HSD. We evaluated the three clusters according to the characteristics on which they significantly differed. Means and standard deviations for each cluster are presented in Table 14. Graphs showing significant differences between cluster means are presented in Figure 5.

Clusters are described individually below. On the whole, they fell along a continuum of behaviors, with Low Compliers demonstrating the least compliance and most empathy, High Compliers demonstrating the most compliance and least empathy, and Moderate Compliers falling in the middle on these markers. High Compliers expressed the most guilt of the clusters, while Low Compliers expressed the least guilt; this lack of guilt in Low Compliers despite their high levels of empathy may be due to the fact that Low Compliers were the most likely to refuse

to do the task, and thus had no reason to feel guilt afterwards, as they had not performed a transgression. On the other hand, High Compliers were the most likely to tear the photo, and still demonstrated substantial guilt despite their low levels of expressed empathy. There were no differences between clusters on observer-rated sociability across laboratory tasks, suggesting that the children's differing patterns of behavior in the Picture Tearing task cannot simply be accounted for by their average levels of social engagement.

Cluster 1: Moderate compliers. Members of this cluster comprised 24.9% of the sample (n = 67). On most behaviors of interest, members of cluster 1 fell between clusters 2 and 3, and were thus moderate in their behaviors. Cluster 1 fell between the High and Low Compliers on expressions of defiance in response to the experimenter's prompt, but only expressed more empathy and hesitation than the High Compliers. With respect to compliance, cluster 1 also fell in the middle on latency to comply, absolute compliance, and extent of compliance, as well as enjoyment in tearing. Finally, cluster 1 also fell into the middle with respect to guilt and relief for the copy.

Cluster 2: Low compliers. Members of this cluster comprised 16.0% of the sample (n = 43). Compared to the other clusters, cluster 2 demonstrated the most defiance. Members showed more empathy and hesitation than the High Compliers, but did not differ from the Moderate Compliers. With respect to compliance, cluster 2 members took the longest time to tear, were least likely to comply, complied to a lesser extent, and showed less enjoyment while tearing than the other clusters. Interestingly, when they did comply, they also showed the least guilt and the least relief when the copy was presented.

Cluster 3: High compliers. Members of this cluster comprised 59.1% of the sample (n = 159). On the whole, members of cluster 3 were the most enthusiastically compliant and the least

empathic and defiant. Compared to the other clusters, members of cluster 3 expressed the least empathy, hesitation, and defiance in response to the experimenter's prompt. In keeping with their lack of verbal dissent, they were also the most likely to comply, to the greatest extent, in the shortest amount of time, and with the most enjoyment. Finally, they expressed the most guilt and the most relief for the copy. Compared to the Low Compliers, they were more likely to respond to the victim by blaming another, lying, or being silent.

External validation. After we determined the relative standing of clusters on the variables used to derive them, we examined whether the clusters demonstrated differential relationships with external variables on which we would expect them to show meaningful differences. We compared each pair of clusters across the external criterion variables utilizing analysis of variance (ANOVA) and Tukey HSD post hoc tests in SPSS (IBM version 22.0).

To determine whether collapsing clusters 2 (Low Compliers) and 4 (which was included in the description of Low Compliers above) substantially changed their relationships with external criterion variables, we compared results of ANOVA analyses between the four cluster solutions and the three cluster solution (with clusters 2 and 4 collapsed). Results showed that the three cluster solution showed the same pattern of significant external associations as the four cluster solution (Child age, Mother-reported attention shifting, Mother-reported shyness, Emotion recognition scores), but showed additional significant between-cluster differences that were obscured in the four cluster solution. Given the lack of differentiation between clusters 2 and 4 on both internal and external criteria, we collapsed these clusters and treated them as one cluster for remaining analyses. As an additional check, we computed the k-means analysis again setting k equal to 3. In this solution, clusters 1 and 3 were identical to their counterparts in the four-cluster solution; the only cases that changed membership were those in cluster 4, which all

were sorted into cluster 2. This check provided further support that clusters 2 and 4 were not dissimilar enough to be analyzed separately. In addition, the three cluster solution is more parsimonious, as there is no theoretical basis to assume that groups differing only on latency to tear and relief would demonstrate substantively different relationships with important correlates.

ANOVA results for the three cluster solution are described in more detail below.

Demographics. Cluster membership was not significantly predicted by child sex or parent ethnicity (Hypothesis 2.7, 2.8, 2.9). Cluster membership was significantly predicted by child age (F = 5.01, p < .01). Post-hoc tests indicate that Moderate and Low Compliers did not differ on age, and Low and High compliers did not differ on age. On average, High Compliers were significantly older than Moderate Compliers (d = 0.47). This is consistent with our correlational findings, given that age was positively associated with compliance and the Higher Compliers demonstrated the most compliance of the three groups (Hypothesis 2.4). It further supports that our hypothesis, that older children would feel more empathy and thus comply less, was incorrect; the press to comply with authority trumped the press to act empathically. However, the fact that High and Low Compliers did not differ on age suggests that factors other than age were contributing to the greater noncompliance and empathy of the Low Compliers.

IQ and emotional intelligence. Cluster membership was not significantly predicted by child verbal intelligence, as measured by the child's baseline PPVT score (F = 826.17, p = .17), contrary to our hypotheses that differences in empathy and anxious arousal would be associated with differences in verbal intelligence (Hypothesis 2.12, 2.13). Cluster membership was also not predicted by emotion recognition scores, indicating that children in different clusters did not differ, on average, in the number of emotions they correctly identified (F = 2.80, p = .53),

contrary to our hypotheses that differences in empathy and anxious arousal would be associated with differences in verbal intelligence (Hypothesis 2.10, 2.11).

Temperament traits. Cluster membership was not predicted by either mother-reported or father-reported traits of Anger, Inhibitory Control, Approach, Smiling, Low intensity pleasure, Impulsivity, Sootheability, Attention Focusing, Sadness, or Fearfulness (Hypothesis 2.14 through 2.29). This is surprising, given that we expected that Fearfulness and facets of effortful control would be related to task compliance. However, clusters did differ on mother-reported Attentional Shifting (F = 8.41, p = .02), such that the Low Compliers demonstrated, on average, poorer ability to shift their attention according to task demands than did High Compliers (d =0.52). Clusters differed on both mother-reported (F = 5.22, p < .01) and father-reported shyness (F = 3.36, p = .04). According to mother-report, High Compliers were the least shy of the sample, as compared to both Low Compliers (d = 0.50) and Moderate Compliers (d = 0.41). For father-reported shyness, differences were no longer significant after post-hoc tests correcting for multiple comparisons. This association was contrary to our hypotheses, as we expected that more assertive children would be less likely to comply. Finally, clusters differed on both mother-reported (F = 3.33, p = .04) and father-reported (F = 3.27, p = .04) High Intensity Pleasure. High Compliers were rated as seeking more intensely stimulating, pleasurable activities than Low Compliers by both mothers (d = 0.46) and fathers (d = 0.50). This finding was consistent with the fact that High Compliers were rated the highest on positive affect of all three groups. It was also inconsistent with our predictions; given that stimulation seeking is a characteristic associated with externalizing behaviors, we expected that children who scored high on High Intensity Pleasure would be less compliant than children who scored low. It is more consistent with our contention that High Intensity Pleasure, as an associated feature of

externalizing behavior, might predict low empathy. Indeed, the High Compliers, who were rated highest on High Intensity Pleasure, also demonstrated the lowest empathy. In this case, the "thrill" of tearing the picture might have been more salient to these children than the opportunity to be noncompliant.

Overall, the most compliant children were also the most stimulation seeking and demonstrated the most effortful control (specifically ability to shift attention), while the least compliant children demonstrated the least effortful control. The most compliant children were also the least shy, suggesting that perhaps they were more motivated than more shy children to acquiesce to the experimenter's request, as a way of promoting the relationship between themselves and the experimenter.

Problem behaviors. Cluster membership was not predicted by any of concurrent problem behaviors measured by the CBCL, including global externalizing and internalizing and CU traits. This result is contrary to our hypotheses that either task compliance in the absence of empathic statements, or task noncompliance, would both be related to externalizing behaviors (Hypothesis 2.1, 2.2, 2.3).

Parenting behaviors. Cluster membership was not predicted by any parenting behaviors measured by the APQ, rated by either mothers or fathers. This is contrary to our hypotheses that harsh and inconsistent parenting would be indirectly related to task compliance (Hypothesis 2.30), given that this type of parenting is empirically associated with less empathy.

Overall. Overall, the most compliant children were also the most stimulation seeking and demonstrated the most effortful control, while the least compliant children demonstrated the least effortful control. These differences in clusters cannot simply be accounted for by age or emotion recognition abilities, as the two extreme groups (High and Low Compliers) did not differ on age

or emotion recognition. There were also no differences in externalizing problem behaviors or in parenting behaviors theoretically linked to problem behavior. Temperament, specifically effortful control, seems to be playing a role independent of cognitive maturity (indexed by age, verbal ability, and emotion recognition) in predicting compliance and empathy in this task. These analyses demonstrate that the coding scheme does allow us to observe meaningful differences in patterns of child behavior within this task.

Gender of Victim as a Predictor of Child Task Behaviors

In order to determine whether the gender of the victim was associated with any differences in child behavior, we first coded each video for gender of the victim (male = 0, female = 1). The vast majority were female (n = 266, 98.5% of videos), with only four male victims (n = 4, 1.5% of videos). We then conducted t tests comparing the means of all of our task variables separated by epoch, task variable totals, the five empirically derived factors, and cluster membership, to determine whether these means differed between male and female victims. Children in the presence of female victims demonstrated more Hunched Shoulders, Social Referencing, and Laughter in epoch 1 and more Hunched Shoulders in epoch 3. These differences, though statistically significant, are theoretically minor. No differences were observed on key constructs (i.e. Empathy, Guilt, any of the compliance variables) related to moral decision making or to any of our hypotheses.

It is also important to note that the male experimenter was the same person in each of the 4 instances of a male victim, so it is unclear whether effects can be generalized to male victims in general. In addition, given that a male victim was present in only 4 videos, and that the base rates of each task behavior were largely positively skewed, these differences are likely due to

chance rather than meaningful disparities. Given these considerations, any meaningful conclusions on the influence of victim sex at this point would be premature.

Prediction of Concurrent Problem Behaviors Using Composites Derived from Factor Analysis

Correlations between independent variables and dependent outcomes.

Picture tearing composites. We conducted correlations between our outcome variables (EXT, INT, CD symptoms, ODD symptoms, Rule-breaking, Aggression, and CU traits) and our independent variables. As independent variables, we included the five composites formed by the factor analysis (Guilt, Appeal to Authority, Noncompliance, Positive Affect,

Avoidance/Withdrawal.

We found few associations between task behaviors and problem behaviors (see Table 15). However, opposite to our expectations, we did find that the Guilt composite was positively associated with mother-reported Aggression (r = 0.15), ODD symptoms (r = 0.17), and EXT (r = 0.14).

Demographic variables. We examined the correlations between problem behavior outcomes and child sex and age (Table 15). With respect to problem behaviors, child sex did not correlate with any of the outcome variables and was excluded from further analyses. This is somewhat unexpected, given that boys typically demonstrate more externalizing behaviors than girls (e.g. Broidy et al., 2003). However, our sample is composed of younger children, mostly of preschool age; some studies have found that sex differences in externalizing behaviors are smaller or nonexistent in the early preschool years (Keenan & Shaw, 1994; Rose et al., 1989). Age demonstrated small positive correlations with CD symptoms (r = 0.17), INT (r = 0.26), and

Rule-breaking (r = 0.16) (mother-reported outcomes only) and was included in block 1 of these regressions.

Verbal and emotional ability. We also examined correlations between problem behaviors and verbal ability (measured by the PPVT overall score) and emotion recognition ability (measured by performance on the emotion training lab task) (Table 15). PPVT score showed a small positive correlation with mother-reported INT (r = 0.16). Of the emotion recognition variables, total emotions correctly identified and misidentifying happy, sad, fearful, and neutral expressions did not correlate with any outcomes. Misidentifying angry expressions correlated positively with CU symptoms (r = 0.29), ODD symptoms (r = 0.18), and Aggression (r = 0.20) (father-reported outcomes only).

Temperament traits. We also conducted correlations between problem behavior variables and higher order temperament traits. We examined correlations between these traits and problem behaviors in order to determine whether to include them as steps in hierarchical regressions according to our planned analyses. Consistent with previous literature, higher Effortful Control (bother father- and mother-reported) was associated with fewer externalizing problems of any kind (mother- or father-reported Aggression, Rule-breaking, ODD, CD, EXT, CU traits), and was not associated with internalizing problems. Mother-reported Negative Emotionality was positively associated with mother-reported Aggression, ODD, INT, and EXT, and father-reported Negative Emotionality was positively associated with mother-reported Positive Emotionality was positively associated with mother-reported Positive Emotionality was positively associated with father-reported Aggression, Rule-breaking, ODD, CD, and EXT, while father-reported Positive Emotionality was positively associated with father-reported Aggression, Rule-breaking, ODD, CD, EXT, or DD, CD, EXT, is positively associated with mother-reported Aggression, ODD, INT, and EXT, and father-reported Negative Emotionality was positively associated with all father-reported Positive Emotionality was positively associated with mother-reported Aggression, Rule-breaking, ODD, CD, and EXT, while father-reported Positive Emotionality was positively associated with father-reported Aggression, Rule-breaking, ODD, CD, EXT, and CU traits.

With respect to temperament traits coded from other laboratory tasks, we examined correlations between fearfulness and compliance and problem behaviors. Fearfulness was not correlated with any of the problem behaviors. Compliance was negatively associated with father-reported Aggression, Rule-breaking, ODD, CDD, and EXT, but did not demonstrate any correlations with mother-reported problem behavior.

Parenting characteristics. Finally, we examined the associations between parenting characteristics rated by mothers and fathers and problem behaviors (Table 15). Father-rated Poor Monitoring, Corporal Punishment, and Appropriate Punishment did not correlate with any of the outcomes (with the exception of a modest positive association between Corporal Punishment and father-rated CU traits, r = 0.20). This is somewhat surprising, given the well-validated association between harsh and inconsistent parenting and externalizing problems. However, consistent with hypotheses, mother- and father-reported Involvement and Positive Parenting were associated with fewer problem behaviors overall, while mother-rated Inconsistent Discipline, Poor Monitoring, and Corporal Punishment were associated with more problem behaviors. Unexpectedly, mother-reported Appropriate Punishment was actually associated with more mother-rated Aggression, ODD, and Rule-breaking, but with less father-reported total INT and CU traits. It is possible that mothers whose children demonstrate more externalizing behavior also view themselves as delivering more punishment in general than mothers whose children are low on externalizing.

Overall. Overall, few associations were observed between externalizing outcomes and Picture Tearing composities. Notably, Guilt was positively associated with aggression, ODD symptoms, and CU traits. Age was positively associated with CD symptoms, INT, and Rule-breaking. Verbal abilities were only correlated with slightly more INT, while overall emotion

recognition did not predict any problem behaviors. Only misidentifying anger specifically was correlated with more CU symptoms, ODD symptoms, and aggression.

As expected, higher effortful control was consistently associated with fewer externalizing problems of any kind, while both higher negative emotionality and positive emotionality were consistently associated with more externalizing problems. Laboratory-observed fearfulness was not associated with any problem behaviors, but laboratory-observed compliance was associated with less father-reported externalizing.

Finally, dimensions of harsh or inconsistent parenting were related to externalizing behaviors, but only for mother-reported parenting. Positive parenting and involvement both predicted fewer problem behaviors.

Hierarchical regressions predicting problem behaviors from task behavior composites. According to our analytic plan, we first performed a series of hierarchical regressions controlling only for child age in block one (where age demonstrated a significant zero order correlation with the dependent variable of interest), with the composite task variables in block two. For each significant regression, we calculated ΔR^2 for each block to determine whether the predictors accounted for significant variance beyond the effects of age and sex.

With respect to mother-reported child outcomes, the laboratory composite variables significantly predicted ODD symptoms over and above child age and sex ($\Delta R^2 = .06$, p = .04, *Cohen's f*² = .06). Only the effect for Guilt was uniquely significant (B = .12, p = .01), but in the opposite direction expected (Hypothesis 2.3). No models predicting father-reported outcomes were significant.

Hierarchical regressions controlling for other child and environment characteristics. Next, we retained the significant regressions and added additional steps

accounting for other child and environmental characteristics. Only variables that had significant zero order correlations with the dependent variable, as described above, were included in each regression. Two regressions were performed for each problem behavior, one in which temperament traits were entered in the block before parenting variables, and one with the order reversed.

In predicting mother-reported ODD symptoms, the effect of Guilt was significant (B = .11, p < .01), but did not account for additional variance beyond the effects of temperament and parenting variables ($\Delta R^2 = .04, p = .08, Cohen sf^2 = .04$) (Hypothesis 2.3). The values of the slope for Guilt and change in R^2 for that block were identical whether parenting variables were entered before or after temperament traits. Thus, regardless of whether parenting characteristics or temperament traits are casually prior in the prediction of ODD symptoms, there is no remaining variance in ODD that is significantly explainable by the addition of Guilt.

Summary of hierarchical regressions using task composites. Overall, there were few significant regressions. We had hypothesized that noncompliance would predict more EXT (Hypothesis 2.1), while anxious arousal/ guilt and empathy would predict less EXT (Hypothesis 2.2, 2.3). In fact, the Guilt composite, which contained both overall guilt and empathic statements, predicted *more* ODD symptoms, but was no longer a significant predictor once variance due to child age, temperament, cognitive characteristics, and parenting was accounted for. Thus, task composites do not demonstrate utility in predicting concurrent externalizing behaviors. It seems that the "slice" of behavior observable in Picture Tearing is not eliciting either the degree or variability of moral emotions that would be indicative of variability in externalizing problems outside of the laboratory.

Prediction of Concurrent Problem Behaviors Using Individual Task Variables

Correlations between independent variables and dependent outcomes. As outlined in our aims, we also investigated the association between single variables derived from the coding scheme and problem behavior outcomes. We utilized the variables indexing total behavior across epochs (Empathy, Hesitation, Defiance, Social Reference, Laughing, Smiling), the variables indexing compliance and extent of compliance (Latency to Tear, Tear or Not, How Much Torn, Enjoyment in Tearing), and the variables indexing guilt and arousal in response to the victim (Gaze Avoidance, Overall Guilt, Relief for Copy). Overall, these variables broadly cover our domains of interest: compliance, empathy, and guilt/anxious arousal.

Similar to our results using factor composites, we found unexpected positive associations between Empathy and mother-reported Aggression, ODD symptoms, and EXT (Hypothesis 2.3). We also found unexpected positive associations between Social Referencing and motherreported Aggression, Rule-breaking, CD, and EXT (Hypothesis 2.2). Finally, we found, as expected, a modest positive correlation between Enjoyment in Tearing the picture and motherreported CD symptoms.

Hierarchical regressions predicting problem behaviors from task behavior.

According to our analytic plan, we first performed a series of hierarchical regressions controlling only for child sex and age in block one, with the composite task variables in block two. For each significant regression, we calculated ΔR^2 for each block to determine whether the individual task behaviors accounted for significant variance beyond the effects of age and sex.

With respect to mother-reported outcomes, total Empathy (B = .08, p = .02) and Social Referencing (B = .05, p = .12) predicted Aggression above and beyond sex and age ($\Delta R^2 = .04$, p = .06, *Cohen's* $f^2 = .04$), with only the effect of Empathy being significant. Total Empathy (B = .06) .17, p < .01) also significantly predicted ODD symptoms ($\Delta R^2 = .05$, p < .01, *Cohen's* $f^2 = .05$). Finally, the combination of Empathy (B = .05, p = .06) and Social Referencing (B = .04, p = .08) predicted EXT over and above age and sex ($\Delta R^2 = .03$, p = .03, *Cohen's* $f^2 = .03$). All effects were small in magnitude. No regressions were significant for father-reported outcomes. Again, the positive association between Empathy and EXT was contrary to our hypotheses (Hypothesis 2.3).

Hierarchical regressions controlling for other child and environment

characteristics. Next, we retained the significant regressions and added additional steps accounting for other child and environmental characteristics. Only variables that had significant zero order correlations with the dependent variable were included in each regression. Two regressions were performed for each outcome, one in which temperament traits were entered in the block before parenting variables, and one with the order reversed.

With the addition of those cognitive, temperament, and parenting variables that were significantly correlated with ODD symptoms, picture tearing task variables only remained significant in predicting mother-reported ODD symptoms ($\Delta R^2 = .03, p = .01$), with total Empathy as the significant predictor of more ODD symptoms (B = .14, p = .01) (Hypothesis 2.3).

Summary of hierarchical regressions using individual task variables. As with regressions using task composites, few regressions were significant. However, notably, total empathic statements continued to predict ODD symptoms beyond the variance explained by cognitive characteristics, temperament, and parenting. This is contrary to our hypothesis that empathic statements would be associated with less EXT (Hypothesis 2.3). It is also different from results for the Guilt composite; once other variables were added, Guilt did not continue to predict ODD symptoms. Guilt contains both overall guilt and empathic statements. These

results suggest that the prediction of ODD symptoms is being driven by empathic statements, and the addition of guilt actually obscures the predictive relationship between empathic statements and ODD symptoms. With respect to our aim of using task behaviors to predict externalizing behaviors, it seems that only empathic statements are a useful predictor of externalizing behavior; however, contrary to our hypotheses, they are a better index of oppositionality than of adaptive moral understanding (and thus *less* externalizing). Noncompliance and anxious arousal within the task did not emerge as useful predictors of EXT in either direction (Hypothesis 2.1, 2.2). Thus, it would appear that while the Picture Tearing task does allow for the elicitation of moral behavior, and variations in this moral behavior between children do not reflect variation in externalizing behaviors outside of the laboratory.

Prediction of Concurrent Temperament Traits Using Picture Tearing Variables

Correlations between independent variables and dependent outcomes. Finally, we decided to conduct a series of analyses examining the associations between picture tearing variables and relevant temperament traits. These associations were not initially outlined in our aims, as we were more interested in predicting externalizing behaviors. However, we observed through our analyses of clusters that clusters differing on empathy and compliance also demonstrated meaningful differences on temperament traits, particularly those related to effortful control (e.g. inhibitory, impulsivity). Furthermore, parent-reported temperament was associated with parent-reported behavior problems, suggesting the possibility that, at the age of children in our sample, picture tearing behaviors are associated with traits related to risk for psychopathology more strongly than behavior problems themselves. Thus, we wanted to assess more stringently how behaviors during the picture tearing task relate to variations in child temperament. Results of correlations are presented in Table 16.

Consistent with results from the cluster analysis, most of the significant associations were observed with facets of effortful control. Our hypotheses for the relationship between effortful control and compliance largely held (Hypothesis 2.27), while the results for effortful control and empathy and guilt violated our expectations (Hypothesis 2.28, 2.29). Specifically, we hypothesized that effortful control would be positively associated with compliance, empathy, and guilt. Broadly, facets of effortful control were indeed modestly related to less overall Noncompliance, less delay in complying (Latency), and greater extent of compliance (How Much Torn), but was unexpectedly related to less Empathy and Guilt (either Overall Guilt or the Guilt Composite). At the facet level, however, the relationships with Empathy and Guilt were significant only for Attentional Shifting and Impulsivity, and Inhibitory Control did not relate to Empathy, Guilt, or Noncompliance at all. Notably, these associations were all small in magnitude and most were observed only for mother-reported temperament traits.

On the whole, the results for inhibitory control are somewhat surprising, given that inhibitory control has empirically and theoretically been associated with better ability to inhibit preponent responses, such as defying an adult authority. Thus, we would have expected that children with better inhibitory control would be less noncompliant, whereas the association in our data was null. However, it is more surprising that impulsivity was positively related to empathy and guilt, while attentional shifting was negatively related to these constructs. Past literature has linked poor effortful control with poorer moral internalization, which would suggest the opposite associations.

With respect to higher order temperament variables, mother-reported Effortful Control was associated with less Noncompliance (r = -0.18) and Defiance (r = -0.14) and father-reported EC was associated with less Empathy (r = -0.17). Mother-reported Negative Emotionality was

modestly associated with more Hesitation (r = 0.14), less absolute compliance (Tear or Not) (r = -0.19), and less extent of compliance (How Much Torn) (r = -0.16). It appears that children higher on negative emotionality were more tentative in their response to the main experimenter's request, potentially because they are more motivated than children low on NE to avoid committing an act that will "get them into trouble" and subsequently cause them to feel additional negative emotions.

Finally, mother-reported Positive Emotionality was modestly associated with more Guilt (composite) (r = 0.18), Authority (r = 0.14), and Relief for the copy (r = 0.17) and father-reported PE was associated with more Social Referencing (r = 0.18). Given that positive emotionality is generally correlated with higher sociability, it appears that children higher on positive emotionality in our sample were accordingly more likely to seek closeness and positive interaction with the experimenters, whether they expressed this by expressing concern for the victim, seeking approval from the main experimenter, or sharing in the victim's positive emotions when the loss of the cherished photograph is repaired.

Hierarchical regressions predicting temperament traits from task behavior. We first performed a series of hierarchical regressions controlling only for child sex and age in block one (where age and sex were significantly correlated with the outcome in question), with the composite task variables in block two. We also performed a series of regressions with age or sex in block one and dummy coded contrasts comparing clusters in block two. For each significant regression, we calculated ΔR^2 for each block to determine whether the predictors accounted for significant variance beyond the effects of age and sex.

The first set of regressions used the five empirically-derived picture tearing factors as predictors in one block. With respect to mother-reported traits, the empirically derived factors

significantly predicted Attentional Shifting ($\Delta R^2 = .09, p < .01$); specifically, children who were more likely to seek approval from the experimenter demonstrated less attentional shifting (Authority; B = .20, p < .05), while children who withdrew through silence or avoided gaze (Avoidance; B = .20, p < .05) were higher on attentional shifting. Impulsivity was significantly predicted ($\Delta R^2 = .12, p < .01$) by more Guilt (B = .22, p < .01) and more appeals to Authority (B= .20, p = .01). On the whole, these results are consistent with the notion that children who have poorer effortful control (here, attentional focusing and shifting) are less equipped to execute and comply with commands issued by authorities. However, another facet of effortful control, impulsivity, was associated not with noncompliance, but with exhibitions of empathy/guilt and anxious arousal. It is possible that, in our task, impulsive children were more likely to comply with the task because the thought of doing something usually forbidden was highly appealing, without considering the consequences. Subsequently, they might be more likely to express signs of guilt once faced with the consequences.

The regression predicting shyness was also significant ($\Delta R^2 = .08, p < .01$), with Noncompliance (B = .49, p < .01) unexpectedly associated with more shyness. This finding may reflect the way that shyness is measured via the Child Behavior Questionnaire. This subscale contains items suggesting a child who is slow to engage with strangers, inhibited around unfamiliar people, and nervous in the presence of novel adults. It is possible that the shyness here reflects an unwillingness or reluctance on the part of the child to engage with the unfamiliar experimenter, rather than simply the opposite of assertiveness/surgency. Thus, the shy child might simply fail to comply because their tendency is to withdraw from the stranger and her request, not because they are being intentionally defiant. Variance in High Intensity Pleasure was also primarily driven by Noncompliance (B = -.32, p = .02), such that children who complied less were rated as being more likely to seek highly stimulating activities; this association is consistent with our hypotheses (Hypothesis 2.18), given that both noncompliant behavior and sensation-seeking behavior are consistently empirically associated with externalizing pathology and with each other.

Variance in Attentional Focusing ($\Delta R^2 = .05$, p = .04) was also predicted by Noncompliance (B = -.33, p < .01), such that more noncompliant children demonstrated worse attentional focusing, consistent with our expectations (Hypothesis 2.27). Higher-order Positive Emotionality ($\Delta R^2 = .06$, p = .02) was predicted primarily by Guilt (B = .10, p < .01), with children who exhibited more guilt also being rated as generally exhibiting more PE. For fatherreported temperament traits, none of the regressions predicted significant change in R² after accounting for age. We further examined the potential effect of cluster membership on temperament. With respect to mother-reported temperament traits, the contrast between High and Low Compliers (B = -.23, p = .02) predicted significant change in the variance of Attentional Shifting ($\Delta R^2 = .04$, p = .02). The contrast between clusters High and Low Compliers (B = .40, p<.01) significantly predicted Shyness ($\Delta R^2 = .05$, p = .01). The contrast between clusters High and Low Compliers (B = -.19, p = .04) also predicted High Intensity Pleasure ($\Delta R^2 = .03$, p < .04) .05). With respect to father-reported temperament traits, the contrast between clusters High and Low Compliers (B = .37, p = .01) predicted Shyness ($\Delta R^2 = .05$, p = .04). These results provide further support for the conclusions suggested by the regressions utilizing factor composites as predictors. Attentional Shifting, Shyness, and High Intensity Pleasure were all associated with membership in the cluster whose members were most noncompliant.

Finally, we examined variance in child temperament measured in the other laboratory tasks. Picture tearing composites did not predict variance in fearfulness in other tasks ($\Delta R^2 = .02, p = .58$), contrary to our expectations (Hypothesis 2.24, 2.25, 2.26). However, they did predict variance in compliance in other tasks above and beyond age ($\Delta R^2 = .05, p < .05$). The effect for Noncompliance (B = -.16, p < .05) was significant, such that noncompliance in Picture Tearing was associated with slightly less compliance in other laboratory tasks, which conforms to our expectations (Hypothesis 2.24). These results suggest that children who were likely to be defiant in general were, in fact, somewhat more defiant in Picture Tearing as well. However, the very small magnitude of this effect is interesting in that it suggests that noncompliance in Picture Tearing and in other tasks were only modestly related to one another. This further suggests that, conceptually, noncompliance in Picture Tearing does not simply reflect a general tendency not to comply, but is being driven by other factors.

Summary of regressions predicting temperament traits from task behaviors.

Overall, task behaviors were most useful in predicting variations in effortful control, though the direction of the associations was opposite to what we had anticipated for empathy and anxious arousal (Hypothesis 2.28, 2.29). Impulsivity was predicted by higher Guilt, while Attentional Shifting was predicted by both less Appeal to Authority figures and more Avoidance/Withdrawal behaviors. Better Attentional Focusing was predicted by less noncompliance, providing further support for a positive relationship between task compliance and effortful control (Hypothesis 2.27). Noncompliance also predicted both more shyness and more high intensity pleasure seeking.

With respect to emotional expression, greater Positive Emotionality was predicted by higher Guilt. This association may be due in part to a link between extraversion / lack of shyness

and a tendency to verbalize empathic statements, as the Guilt composite also contains empathic statements. There were no significant regressions for Negative Emotionality.

With respect to cluster membership, the difference between High and Low Compliers predicted variance in Attentional Shifting, Shyness, and High Intensity Pleasure. This again supports that ability to shift attention allowed for less noncompliance, while children high on stimulation seeking and low on shyness were more likely to perform the task, perhaps either because it seemed fun or because they wanted to please the experimenter.

With respect to performance in other laboratory tasks, there were no significant regressions predicting observed fearfulness (Hypothesis 2.24, 2.25, 2.26). Noncompliance in Picture Tearing was associated with lower compliance in other laboratory tasks (Hypothesis 2.24). However, the modest magnitude of the association suggests that noncompliance in Picture Tearing was not wholly reflective of a general tendency not to comply, but was also being driven by other factors unique to the task.

These data suggest that, while Picture Tearing behaviors have limited utility in predicting variation in child externalizing behaviors, they are indeed reflective of variations in important aspects of child temperament. In particular, variations in child noncompliance, empathy, and anxious arousal/guilt are associated with differences in effortful control faculties, such that children highest on effortful control were the most compliant and expressed the fewest reservations in response to the experimenter's request. Task behavior within Picture Tearing seems to reflect a child's ability to follow a request and to inhibit task-inconsistent behavior.

Longitudinal Prediction of Problem Behaviors Using Composites Derived from Factor Analysis

We conducted growth curve analyses looking at effects of our composite picture tearing variables and cluster membership on problem behavior outcomes (Aggression, Rule breaking, CD, ODD, CU traits, EXT, and INT) over time, from baseline out to the 24-month assessment, across all 6 waves. Multilevel models (MLM) were estimated using HLM 7.0 (Scientific Software International Inc.). Repeated assessments (CBCL scales) were nested within participants at level-1 of the MLM; child-level factors (i.e., picture tearing variables) were entered at level-2 of the models. For each outcome, the level 1 model describes the linear and quadratic effects of age on the outcome. The intercept was set at 36 months (the youngest age of any participant) and age-related change was modeled as change in months after that. We used MLM to test whether picture tearing task variables at level-2 predicted individual differences in children's initial level (intercept) and age-related change in behavior problems.

At level 1, CD symptoms increased modestly along a quadratic trajectory as children aged (p = .04), while ODD symptoms and total EXT did not demonstrate significant age-related change without the addition of level 2 predictors, as described below. Total INT did not demonstrate age related change in the sample, nor did picture tearing variables significantly affect change in INT over time.

With the addition of level 2 predictors (picture tearing variables), we observed that individual differences in intercepts and age-related trajectories for CD, ODD, and EXT demonstrated significant associations only with Guilt. Unexpectedly, children high on Guilt started out at baseline with higher levels of EXT, CD symptoms, and ODD symptoms than their low-Guilt peers (Hypothesis 3.2, 3.3). However, these children also decreased more quickly

over time than low-Guilt children in EXT, CD symptoms, and ODD symptoms. This decrease *is* consistent with our assertion that empathy and anxious arousal/guilt would negatively predict EXT (Hypothesis 3.2, 3.3). Specifically, at the time of laboratory assessment, children who expressed more empathy and guilt were also higher than their peers on externalizing pathology. However, this effect appears to be limited to concurrent associations. When considering the data longitudinally, high levels of Guilt appear to predict a lessening of externalizing pathology and a return to a more adaptive developmental trajectory over time.

Considering the data as a whole, this may reflect an intersection of processes, such that children who are more impulsive and show less effortful control at baseline display more guilt and express more empathy *at the time of the laboratory task*. High levels of impulsivity and low levels of effortful control might also dispose them to be rated by their parents as displaying more externalizing pathology, again at the time of baseline assessment. However, their high levels of empathic concern and guilt could potentially buffer against the increase of externalizing pathology in these children as they age. On the other hand, children who demonstrate low levels of guilt at baseline do not demonstrate a decrease in externalizing pathology as they age. It is possible that these children are already disposed to be on a trajectory of consistent EXT over time, and their low levels of guilt at the baseline assessment are an early sign of this. Put another way, their absence of guilt might represent a risk factor for stable EXT over time.

In addition, children high on Positive Affect during the picture tearing task showed a modest nonlinear decrease of total EXT over time compared to children low on Positive Affect, suggesting that positive affect could also act as a protective factor against the increase of externalizing pathology as children age. Further investigation of this possibility is outside the scope of this particular study, as positive affect is not unique to moral decision making.

Summary of longitudinal prediction of problem behaviors. Overall, there were few significant regressions of longitudinal outcomes, which is consistent with concurrent results. Guilt drove the significant analyses. Children high on Guilt started out at baseline with higher levels of EXT, CD, and ODD than low-Guilt peers, but decreased more quickly over time on these problem behaviors. Thus, children high on Guilt initially appeared to have more problematic behavior, but returned to a more adaptive trajectory over time. This is partially consistent with our assertion that anxious arousal/guilt and empathy would negatively predict long-term EXT (Hypothesis 3.2, 3.3).

We had intended for task behaviors to be useful in predicting changes in externalizing behavior. For the most part, behaviors during Picture Tearing were not useful in predicting change in problem behaviors over time. However, Guilt (including empathic statements and overall guilt) did, in fact, demonstrate utility in predicting an adaptive decrease in externalizing behaviors for children who were initially high on both externalizing and guilt. Thus, it seems that the Guilt composite in our task does reflect moral emotions thought to act as a protective factor against externalizing problems. Guilt behaviors observed within Picture Tearing appear to meaningfully distinguish children who, despite appearing to have more EXT at the time, are more likely than their low-Guilt peers to decrease in their EXT as they age.

DISCUSSION

This study examined and sought to validate a coding scheme for the Picture Tearing laboratory task, with the aim of examining the structure of empathy and compliance behaviors in a naturalistic, morally-relevant scenario. Specifically, we aimed to refine the coding scheme for the Picture Tearing task and create recommendations for its use in future research. In order to examine and refine the coding scheme, we applied it to a sample of children aged 3 to 7 years and examined its performance and utility in several areas: describing variations in moral decision making behaviors among children aged 3 to 7 years, determining the underlying structure of these behaviors, testing whether coherent patterns emerged in these behaviors using personcentered analysis, and determining whether behavior assessed in this task showed concurrent and predictive validity for with child externalizing problems and temperament traits.

On the whole, we found that behaviors observed in Picture Tearing do distinguish children on a number of external markers theoretically important to moral decision making, including age, emotion recognition ability, and aspects of effortful control and surgency. However, the utility in predicting variations in child EXT was somewhat different than we had expected. Guilt and empathy (operationalized as expressed empathic statements) actually predicted more oppositionality in children concurrently. However, these behaviors also predicted a decrease in EXT over time. Thus, the morally-relevant behaviors children exhibit in Picture Tearing do appear to translate outside of the laboratory.

The measurement of empathy, guilt, and compliance in vivo has significant methodological implications for research in moral decision making, which has historically been predicated on hypothetical vignettes or parent report of child behavior. The Picture Tearing task provides the unique opportunity to circumvent problems of reporter bias and to clarify

inconsistencies in past results, which may be attributed to problems with the ecological validity of previous paradigms. It is critical that we understand the nature of moral decision making in children in this age range, across which empathy and moral decision making is first developing and undergoing rapid changes. Insight into how empathy and moral decision making develop and function in young children is a crucial foundation for future work into how these processes can be encouraged or disrupted. While there are important limitations to the utility of our coding scheme for the Picture Tearing task, as we will discuss below, the data on the whole suggest that it is a useful method of examining variations in children's moral behavior; in particular, it provides incremental information above that which can be gleaned from self- and parent-report methods.

Confirmation and Extension of Previous and Hypothesized Findings

In order to determine whether our coding scheme was truly measuring variations in behaviors relevant to moral decision making, we first investigated the patterns of behaviors observed during the task itself. On the whole, our data suggest that the coding scheme measures behaviors that are consistent with theoretical expectations for child behavior in a morally ambiguous scenario. Most importantly, our observational approach allowed us to actually observe behaviors that have, heretofore, been measured only as hypotheticals.

As we had expected, most children complied with the experimenter / authority figure. Moreover, most children complied with little to no resistance. The base rate of all behaviors was positively skewed, with most children complying relatively quickly and without comment. This is consistent with the age of our sample (3 to 7 years old), given that children at this age are typically socialized to accept the commands of authority figures (Kohlberg & Kramer, 1969; Piaget, 1965). It is also consistent with previous research using both neutral (Kochanska, Coy, &

Murray, 2001; Kochanska et al., 1997) and morally ambiguous (Milgram, 1963; Shanab & Yahya, 1977) tasks, wherein both children and adults have been shown to comply with the request of a perceived authority figure, even if the request is putatively harmful to another person. As we had anticipated, children who demonstrated more enjoyment in the task, more social referencing, and more positive affect were also more likely to comply and to do so quickly and thoroughly.

Despite high rates of compliance in general, there was sufficient variation in the extent of compliance to suggest that even children as young as three years of age do not uniformly respond to external strictures without an awareness of the meaning of their actions. Some children waited much longer to comply, required more prompts, and tore the picture less thoroughly than did other children. Early moral development models (Kohlberg & Kramer, 1969; Piaget, 1965) have suggested that children are pragmatic, premoral agents who respond to commands based simply on the dictates of authority figures. In short, these models would suggest that children of the age of those in our study responded to the experimenter's request based on their relative fear of punishment, without considering the meaning of the request in terms of their own personal values or broader issues of morality. Moreover, both Piaget (1965) and Kohlberg (Kohlberg & Kramer, 1969) contended that young children do not experience empathy or guilt, and might have expected to observe signs of these moral emotions only from the oldest participants in the sample (aged seven years) and still only rarely.

However, our data as a whole contradict this view of children as uncritically ruleresponsive and unempathic. A survey of the manner in which children responded to the experimenter's request verbally clearly supports that many children were not only critically evaluating the meaning and appropriateness of the authority figure's request, but were

experiencing the moral emotions that early researchers ascribed exclusively to older children. For example, multiple children interrogated the experimenter about the request ("Are you sure? Did she say you could do this? Why do you want me to?"). Others, even the youngest children in the sample, revealed that they were considering the moral implications of the task and anticipating the victim's feelings ("What if she's mad? She's going to be sad. That's not nice."), while some even suggested a less harmful alternative ("But it's her favorite. How about we rip this other picture instead."). Some children condemned the experimenter for what they clearly perceived as bad behavior ("You're not supposed to do that. You're mean. You're being a bully."), and a small number of children went so far as to threaten the experimenter with punishment ("I'm going to tell her what you said.").

One could argue that the children were simply parroting statements taught to them by parents, teachers, and other socializing agents. Even so, whether or not these statements are indicative of self-chosen moral principles, they *do* reveal that many of the children were not simply weighing the potential for punishment when choosing to comply or not. Instead, they were concerned about the effect of the action on the victim and were weighing the request against a basic set of internalized moral principles. This interpretation is consistent with more current research suggesting that even children as young as 24 months of age demonstrate at least primitive forms of moral emotions and possess a basic awareness of "right" and "wrong," regardless of whether they can articulate these principles in a sophisticated matter (Zahn-Waxler, Radke-Yarrow, Wagner, & Chapman, 1992).

Some children did remain defiant and refuse to comply with the task. This absolute noncompliance was, as we had expected, positively associated with statements reflecting empathy and defiance. However, the fact that most children complied in spite of expressing
reservations does suggest that, even though young children experience empathy and understand basic moral principles, they cannot necessarily translate their empathy into absolute moral behavior (i.e. refusing to tear the picture) when faced with the strong social press to comply with an authority figure. Yet, that does not signify that their moral sensibilities were completely overridden by the pressure of authority. As we had anticipated, statements reflecting empathy, defiance, and hesitation were all positively associated with noncompliance in a *relative* sense. Most of these children tore the picture eventually, but, in proportion with their aforementioned verbalizations, they delayed longer in tearing, required more prompts from the experimenter before starting to tear, and/or tore the picture less extensively (e.g. ripping only a corner of the picture instead of "tearing it up" completely as instructed). The latter behavior suggests a surprisingly sophisticated sense of moral relativism; that is, if the child was going to "have to" listen to the adult authority, then they were also going to mitigate the damage as much as possible.

In further support of the contention that expressions of guilt after a transgression emerge as early as 24 months of age (e.g. Aksan & Kochanska, 2005; Kochanska & Aksan, 2006), all facets of compliance were positively associated with guilt. This effect was observed whether guilt was operationalized as observer ratings of how guilty the child seemed or objective displays of nonverbal discomfort (e.g. hunched shoulders, squirming, avoidance of the victim's gaze), a method by which guilt has been operationalized in some previous research (Kochanska & Aksan, 2006). Moreover, we observed a coherence between the moral emotions; empathic statements and guilt were significantly and positively correlated. In fact, of the three types of verbalizations coded, only empathic statements were associated with guilt, such that children who verbalized more empathy towards the victim exhibited more guilt when the victim returned.

Empathic statements were associated not only with guilt, but also other moral behaviors of interest. Children who demonstrated high expressed empathy for the victim were more likely than their peers to apologize to the victim or attempt to make reparations. This finding represents a theoretical replication of earlier findings that children as young as 24 months are capable of taking the perspective of a person in distress and attempting to mitigate that person's distress (Zahn-Waxler, et al., 1992). While children in this previous study were likely to offer comfort in forms that they themselves might desire (e.g. offering a teddy bear), the children in our sample (who were at least 12 months older than in the previous study) demonstrated more sophisticated perspective taking, in that the reparations they offered were more appropriate to the victim and the situation (e.g. offering to help the victim put the photograph back together).

After observing the associations between task variables, we attempted to determine the underlying structure of these behaviors, via both a variable centered approach (factor analysis) and a person centered approach (cluster analysis). With respect to factor analysis of the behaviors, the results partially confirmed our hypotheses about the underlying structure and partially defied our expectations. As we had expected, a clear factor emerged in all solutions that reflected noncompliance. This factor reflected defiant statements, the number of prompts given, latency to tear, and how much the child tore the picture, which was identical to our hypothesized results.

Additionally, we did observe a coherence between empathic statements and relief for the copy of the photograph, which we had hypothesized would reflect the child's ability to take the victim's perspective. However, empathy and guilt, the latter of which we had thought would reflect arousal, did not separate as cleanly as anticipated. In fact, a five factor solution provided the best fitting model, rather than the three factor solution we had hypothesized. Our

hypothesized arousal factor, in fact, seemed to conflate different aspects of arousal. Whereas we had suggested that smiling and laughter might reflect internal discomfort, these variables cohered only with each other and did not load with the other variables meant to reflect aversive arousal. Indeed, at the level of zero order correlations, both smiling and laughter are modestly to moderately associated with better compliance, are positively associated with other variables reflecting positive affect (Enjoyment, Relief), and are unrelated to variables reflecting discomfort (with the exception of a small negative correlation between laughter and gaze avoidance). Taken as a whole, it appears that this factor reflects genuine positive emotionality rather than nervous smiling or laughter. Finally, Guilt crossloaded onto a factor with Empathy and Relief, whereas we had expected that Guilt would separately reflect physiological arousal while Empathy would reflect perspective taking or affective empathy. This suggests that the behavior measured by our overall guilt code covaries sufficiently strongly with empathy as to load onto one factor. Indeed, it is possible that guilt as observed by our coders reflects, in part, an outward expression of affective empathy. This is consistent with the literature's assertion that guilt results from the internal empathic awareness of another's distress (e.g. Hoffman, 1975; Kochanska & Aksan, 2006). It seems likely that, at the level of coding observational data, parsing these highly correlated phenomena into discrete, observable behaviors is difficult, if not impossible.

The other aspects of arousal were best represented by two factors. One set of variables (Silence, Gaze Avoidance, Overall Guilt) seemed to reflect a withdrawn and avoidant response style. That Guilt crossloads onto this factor suggests that our overall Guilt code reflects both empathic feeling (perhaps affective empathy) and the aversive internal state we initially expected, and which also manifests as withdrawal. The final set of codes (hesitation, blaming the other person) seemed to reflect a tendency to appeal to authority. It is possible that both the

Avoidance and Authority factors reflect an internal state of aversive physiological arousal, in that they both contain behaviors that attempt to modulate this arousal. Importantly, many of the nonverbal discomfort behaviors (Social reference, Hunched shoulders, Squirming) and response behaviors (Blame self, Apologize, Lying) that we expected to load onto an arousal factor did not have substantially large loadings onto any factor and were dropped from analysis. These were also all behaviors with very low base rates of endorsement across videos, so it is possible that there was not enough variance in these variables for them to factor meaningfully.

Our person-centered analysis of the task behaviors was more exploratory, but yielded patterns of behavior that cohered in interpretable ways. Results of this analyses suggest that the Picture Tearing coding scheme yields codes that describe meaningful variations in child behavior, and that these behaviors cohere into patterns that meaningfully characterize distinct groups of children. Analyses suggested that a three cluster solution best fit the data. High Compliers represented the most compliant group; these children were most likely to comply, complied in the shortest time with the fewest number of prompts, and tore the picture to the greatest extent. Low Compliers represented the least compliant group on all compliance variables, with Moderate Compliers falling in the middle on most behaviors. Consistent with our theorized associations between task variables, High Compliers also expressed less empathy and hesitation than Low Compliers and expressed the least defiance in the sample. They also exhibited more smiling and enjoyment in tearing and more discomfort in the form of hunched shoulders than Low Compliers. Upon the victim's return, cluster differences for guilt and relief mirrored those for compliance, with children from the most compliant cluster demonstrating the most guilt and relief and vice versa; again, Moderate Compliers fell in the middle. Similar results were observed for squirming behaviors. There were fewer differences in other response

behaviors, with the least compliant children, Low Compliers, expressing less blame for the other person, less silence, less lying, less hunched shoulders, and less smiling than the most compliant children, the High Compliers. The Low and Moderate Compliers did not differ significantly on these variables.

Rates of cluster membership largely conformed with our expectations for child behaviors. High Compliers, the most compliant group, also contained the majority of participants (n = 159). Consistent with our hypothesis, most children were willing and quick to comply, and subsequently demonstrated guilt when faced with the victim. Low Compliers (n = 43), the least compliant group, contained the fewest participants. On the whole, the cluster data suggest a general coherence between task compliance and outward indications of guilt, and that children who exhibit these behaviors tend not to express empathy or defiance. These data provide further support for our contention that complying with the experimenter in this task is the most normative response for children in this age group.

Furthermore, we did not find differences in cluster membership by child sex, which runs counter to some literature suggesting that girls demonstrate more empathy than boys (e.g. Garaigordobil, 2009; Hoffman, 1977; Kochanska, DeVet, Goldman, Murray, & Putnam, 2008) and compliance (e.g. Kochanska, 2002; Kochanska, Coy, & Murray, 2001; Kochanska, Tjebkes, & Forman, 1998), but is consistent with other literature suggesting that sex differences in empathy are either spurious (e.g. Eisenberg & Lennon, 1983) or do not emerge until middle childhood or adolescence (Rose & Rudolph, 2006), and that there are no sex differences in compliance (e.g. Abe & Izzard, 1999; Braungart-Rieker, Garwood, & Stifter, 1997; Higbee, 2012). Consistent with cluster results, we did not find significant mean differences between boys and girls for most of the individual task variables or factors. Girls exhibited modestly more

squirming in response to the victim and relief in response to the copy of the photo, but there were no significant differences for any other variables, including empathy.

Our failure to observe notable sex differences, where other studies have, may result from a difference in methodology. With respect to empathy, most previous work has used hypothetical vignettes or self-report, and have found that girls are more likely to articulate concern for the victim and to discuss themes of care and compassion (Gilligan & Attanucci, 1988). However, as multiple lines of research have found, the coherence between a person's hypothetical behavior and their actual moral behavior is generally low (e.g. Aksan & Kochanska, 2005; Kochanska, Aksan, & Nichols, 2003; Kochanska, Forman, Aksan, & Dunbar, 2005; Kochanska, Padavich, & Koenig, 1996; Eisenberg et al., 2010; Hartshorne & May, 1928; Zahn-Waxler, Radke-Yarrow, Wagner, & Chapman, 1992). Moreover, our measurement of empathy is predicated on the child verbalizing empathic statements to the experimenter in what the child believes to be an actual dilemma. As supported by the low base rate of empathic statements in our sample, most children seem to be unlikely to express empathy aloud in defiance of the authority figure's request, even if they are feeling it or thinking about it internally. Had we asked the children what they were thinking during the task, we might have observed that girls express more empathy, as other studies have. Our study at present does not allow us to examine this possibility.

While many of the associations between factors or cluster membership and external predictors were unexpected, as we will discuss below, we did observe significant negative associations between Noncompliance, Defiance, and Empathy and overall Effortful Control. In fact, behaviors derived from the Picture Tearing coding scheme seem especially useful in distinguishing children on the basis of their effortful control. The data for Noncompliance and

Defiance are consistent with findings from laboratory tasks using benign (rather than morally aberrant) commands, which have found that better effortful control is also related to better task compliance (Kochanska, Coy, & Murray, 2001; Kochanska et al., 1997). However, the results for empathy are somewhat unexpected. There is some work to suggest that effortful control is linked to high moral internalization and developed conscience (e.g. Kochanska & Knaack, 2003; Kochanska, Murray, & Coy, 1997), which would suggest that we should have observed a positive association between effortful control and empathy. It is possible that effortful control could allow a child to focus on the potential consequences of their actions and to avoid transgressions. Thus, we might have expected that effortful control was positively related to empathy and guilt. However, again, these findings have been predicated on a small number of studies which utilized benign commands.

Given our findings for empathy, previous findings relating EC to moral internalization and conscience may reflect the ability of children with high effortful control to inhibit impulsive or attractive transgressive responses and to instead choose morally-consistent acts. It may be the case that children with high EC seem more empathic because they are better able to focus their attention on the potential consequences of their transgressions, and to avoid transgressing. However, it does not necessarily follow that they *experience* more empathy than their peers, only that they are better able to act on their empathy appropriately. In the case of our study, effortful control appeared to allow children to inhibit their proponent *empathic* response and instead comply with the unempathic action requested by the experimenter, whereas children with poorer effortful control seemed less able to inhibit their protestations. Given our operationalization of empathy as verbalizations, it seems likely that whether children were experiencing empathic thoughts about the victim, children who were more impulsive and less able to focus their

attention were more likely to "blurt out" these thoughts. Our coding scheme seems especially sensitive to morally-relevant behaviors that distinguish children on the basis of their levels of effortful control and its facets.

Our exploration of effortful control was partly exploratory, given that our primary aim was to distinguish children on the basis of externalizing behaviors rather than effortful control. It would be useful in future studies to further explore the utility of the Picture Tearing coding scheme in measuring behaviors that are relevant to a child's effortful control. Here, we found that variations in task behavior predict variations in parent-reported EC. An more stringent test of the notion that Picture Tearing behaviors relate to meaningful differences in EC would be to examine whether task behaviors predict variations in EC observed in other laboratory tasks or assessment specifically designed to measure EC or executive functioning, such as a Go/No Go task. Furthermore, it would be useful to determine whether the association between task variables and variation in EC is limited to concurrent associations, or whether task behaviors are predictive of meaningful variation in EC across time.

Unexpected Findings

We had originally hypothesized that the Picture Tearing coding scheme would be useful in identifying morally-relevant behaviors that would be predictive of variations in child externalizing pathology. In particular, we had expected that children who demonstrated little empathy or guilt in this task would also score higher on measures of externalizing behavior. However, the relationship between task behaviors and externalizing differed greatly from what we had anticipated. The most unexpected finding across analyses was the consistent association between empathy and/or guilt and externalizing pathology. We had predicted that failure to comply might be associated with defiance and oppositionality. This was observed in our results,

such that children who expressed defiance tended also to be rated by their parents as having more symptoms of ODD and other externalizing pathology. We had assumed that, on the other hand, empathic statements would reflect a mature level of perspective taking towards the victim and an understanding that the experimenter's request was morally wrong. Subsequently, we assumed that this mature perspective taking would reflect appropriate moral socialization, as operationalized by a lack of significant externalizing problems.

We had not anticipated, however, that empathic statements might actually reflect defiance and boldness. In effect, empathic statements seemed to be more reflective of an impulsive child who failed to inhibit their verbalizations than a morally mature child. Despite the perspective taking reflected in their statements about the victim's potential mental state, children who scored higher on the Guilt composite (which included empathic statements and overall guilt) were also rated by their parents as having more ODD symptoms, although this association did not hold when parenting and temperament variables associated with ODD symptoms were included in the regression. Furthermore, the association between empathy/guilt and externalizing problems was observed both concurrently and longitudinally, across follow-up periods out as far as 24 months from the laboratory visit. However, we did not find any differences in externalizing pathology between our three clusters, even though they differed significantly on empathy, compliance, and guilt.

In a similar vein, despite robust findings in the literature that poor emotion recognition is associated with low empathy and low anxious arousal (e.g. Blair & Coles, 2000; Blair, Colledge, Murray, Mitchell, 2001; Dadds, Perry, Hawes, et al., 2006; Munoz, 2009; Blair, Budhani, Colledge, Scott, 2005; Fairchild, van Goozen, Calder, et al., 2009), we did not find significant associations between performance on our emotion recognition task and empathy or guilt. Lack

of findings may be attributable to the nature of our emotion variables, which index whether a child has correctly identified an angry, sad, surprised, happy, or fearful face, respectively. There is only one trial per affective face, whereas previous research finding effects for emotion recognition errors has typically employed multiple trials and thus has increased power to find effects.

Furthermore, we did not observe the associations between behavioral responses to the Picture Tearing task and child temperament traits that we had expected. Primarily, we hypothesized that empathy would be associated with temperament traits that are themselves associated with low levels of externalizing problems, and vice versa. However, as with our findings for EXT, our findings for temperament traits suggest that empathy in this task is not purely tapping into perspective taking or affective empathy. We had anticipated that fearlessness, a quality associated with CU traits and thus a lack of empathy and remorse, would be associated with less concern for the victim, lower anxious arousal or guilt, and greater willingness to engage in, and even enjoy, committing an immoral and potentially exciting action. This finding is, however, consistent with our surprising findings that empathy and guilt were related to *more* externalizing problems. Our findings for fearlessness were thus not suggestive of a link between low empathy and high fearlessness. In fact, the absence of empathic responses, low anxious arousal or guilt, and enjoyment in tearing were not associated with fearlessness at all. Furthermore, fearlessness did not distinguish clusters, even though all three clusters differed significantly on their absolute and relative levels of compliance.

Moreover, we did not anticipate the positive relationship between empathic statements and low levels of facets of effortful control. While previous literature does not account for the effects of effortful control on a child's ability to comply with harmful or immoral orders, as most

studies utilize benign commands, it does suggest generally that high effortful control is related positively to both compliance (Kochanska, Coy, & Murray, 2001; Kochanska et al., 1997) and adequate moral internalization, operationalized here as few parent-reported externalizing problems (e.g. Kochanska & Knaack, 2003; Kochanska, Murray, & Coy, 1997). In fact, in our sample, we either observed the opposite or found no effect. When considering our factors and discrete task variables, there was a significant negative effect of total empathy, total guilt, and the Guilt composite on overall effortful control. At the facet level, total empathy, total guilt, and the Guilt composite were related to poorer Attentional Shifting and higher Impulsivity. With respect to clusters, differences between clusters (which differed from each other primarily in terms of compliance and guilt) were not observed for Inhibitory Control or Impulsivity. However, the Low Compliers, who demonstrated the lowest compliance and guilt, also showed poorer ability to shift attention according to situational demands than other children in the sample. This finding, though unexpected, is consistent with findings for task variables. On the whole, it suggests that variation in behaviors during the Picture Tearing task is reflectively primarily of variation in a child's ability to sustain and shift attention and inhibit oppositional or defiant responses. As described above, it would be useful for further research to more stringently test the utility of Picture Tearing behaviors in predicting variation in a child's effortful control faculties.

Given the unexpected nature of these findings, it is important to note that parent report of externalizing problems may not be the cleanest or best operationalization of adequate moral socialization. Indeed, externalizing behaviors might be motivated by a variety of factors other than deficiencies in empathy and moral decision making abilities, such as impulsivity or negative affectivity. Furthermore, the hypothesis that externalizing behavior reflects inadequate moral

socialization, and thus should be associated with low levels of empathy and guilt, is predicated on the assumption that what parents are rating as high externalizing behavior is synonymous with the construct of high externalizing behavior as defined in the literature. In fact, multiple studies suggest that the norms parents consciously or unconsciously utilize when rating their child's own behavior problems are not necessarily consistent with child report of the same behavior, and may be affected by parent characteristics (e.g. Kolko & Kazdin, 1993; Stranger & Lewis, 1993). Teacher report, on the other hand, has been shown to reflect severity of a child's externalizing behavior in a way that is more consistent with the child's "actual" behavior (as operationalized as the validity of teacher ratings of EXT in predicting future behavioral problems, mental health diagnoses, and referrals), potentially because teachers are more familiar with a wide variety of children and are thus better able to compare a given child's behavior to age norms (Stranger & Lewis, 1993). Thus, while parent report provides invaluable information about a child's behavior, it is possible that the positive associations between EXT and guilt we observed would be more modest or nonsignificant if other methods of assessing EXT (e.g. teacher report, observation) were utilized. Indeed, this possibility provides additional support for the use of Picture Tearing and our coding scheme in future research on moral development. Based on the current literature, largely predicated on self- and parent-report, it would be expected that a combination of empathy and noncompliance in this task was reflective of mature moral decision making. However, our study illuminated the possibility that complying with an authority figure in this case is actually the more adaptive choice. Our observational method provided information incremental to what has been found in previous research.

Finally, the use of scales from the CBCL allows us to assess a wide variety of externalizing problems and compare to the wealth of previous ASEBA-based literature, but the

CBCL was not intended as a comprehensive, nuanced measure of externalizing pathology. Furthermore, it contains few items relating to moral emotions, though these are the key pieces of externalizing pathology that we would expect to be related to unempathic behavior in the picture tearing task. Future work would benefit from including dedicated measure of moral emotions, such as the Inventory of Callous-Unemotional Traits (Kimonis, Frick, et al., 2008), as a more stringent test.

Findings for age were also somewhat complex with respect to our hypotheses. While in general we would expect older children to comply more readily with a command, we also considered that older children might show more mature empathy and might, in our particular experiment, comply less. Indeed, past studies of hypothetical vignettes have found that older children, in comparison to younger children, tend to view adult commands for harm to be illegitimate (Damon, 1977; Laupa, 1991, 1994). We did find significant correlations between age and both empathic statements and the Guilt factor, which were further supported by significant differences for empathy and Guilt means between younger children (participants aged 3 to 4 years old) and older children (participants aged 5 to 7 years old). Similar results were observed for hesitation, social referencing, and relief for the copy of the photo.

However, age was also unexpectedly *positively* associated with task compliance. While age was uncorrelated with absolute compliance, it did correlate with facets of compliance, such that older children demonstrated a shorter latency to comply, tore the photo more extensively, and exhibited more enjoyment while tearing the photo. We observed similar results when we examined the data from a person centered perspective. Members of the most compliant cluster, the High Compliers, were significantly older than the Moderate Compliers, who were middling on most of behaviors. High Compliers also demonstrated the most enjoyment in tearing the

photo. However, they did not differ in age from the least compliant group, the Low Compliers, nor did the Low and Moderate Compliers differ in age from each other, suggesting that our findings for age and compliance may vary somewhat depending on a child's other patterns of behavior.

We hypothesized that older children would view the command as illegitimate and that their empathy for the victim would trump the pull to comply with an authority. However, despite expressing more empathy, older children were largely more compliant than younger children, which is consistent with findings from observational tasks using both benign commands (e.g. Braungart-Rieker, Garwood, & Stifter, 1997; Kochanska & Aksan, 1995; Kochanska, Aksan, & Koenig, 1995) and more extreme harmful commands (Shanab & Yahya, 1977). In both cases, older children were more compliant than younger children. On the whole, age appears to be an important predictor of empathy and compliance. However, given that we controlled for age across our analyses, it is not the case that variability in outcomes due to age is masking variability in outcomes due to other child factors. Even controlling for age, we still observe that empathy does not substantively predict noncompliance, and is positively and unexpectedly associated with externalizing pathology.

Conclusions and Areas for Future Study

Validity of the picture tearing coding system and recommendations for future use. On the whole, our data support that the Picture Tearing coding scheme has utility for describing meaningful variations in children's moral decision making behaviors, variations which also reflect variability in their age, cognitive maturity, and temperament. In particular, task behaviors relate to variations in effortful control, while their relationships to externalizing behavior are different from what would be theoretically expected. Thus, results using the observational

coding scheme provide information about the potential meaning and predictive utility of a child's morally relevant behavior, information that is somewhat different from and incremental to the information derived from previous questionnaire-based literature. Use of this coding scheme in future may help to illuminate inconsistencies in the questionnaire-based literature, particularly inconsistencies between a child's empathy as measured by self- or parent-report and a child's actual moral behaviors.

The aforementioned results support the advancement of several recommendations for researchers planning to utilize this task and coding scheme. With consideration of our results, we have created a finalized coding scheme retaining the most useful and theoretically important variables (see Appendix C). Variables that demonstrated few to no important associations in either factor or cluster analysis were removed (i.e. Hiding, Lip Biting, Crossed Arms). We also recommend that, in future, different coders be assigned to code the 1st/2nd epochs from those assigned to code the 3rd epoch. It is possible that the correlations we found between empathic behavior in epochs 1 and 2 and guilt in epoch 3 can be due in part to the fact that coders observed the whole task, and were assuming that children who had expressed the most empathy were also guiltier than their peers. Because we had the same coder code the 3rd epoch in future will ensure that codes are based purely on the child's behavior during that epoch, and do not merely reflect their behavior earlier in the episode.

We also have several recommendations to how the task is conducted that could potentially increase the utility of the final codes. First, it is suggested that the length of epoch 2 be standardized. In our task, most of the epoch 2 codes were very skewed, with the majority of children scoring a 0 for any given behavior. This might be due to the fact that the length of

epoch 2 was as brief as 20 seconds in some video episodes, which did not allow sufficient time for the child to exhibit behaviors. While it is still possible that most children who not exhibit much behavior in this epoch, given that they are alone in a room and are unlikely to speak to themselves, standardizing the epoch at a length of one minute or more might allow additional time for the child to exhibit bodily discomfort.

In addition, it would be useful to ask for self-report data from the child during or at the conclusion of the task. We recommend that the child be asked questions meant to access their thoughts about the task and their behavior: "Why did you tear the picture?" "What were you feeling when you were asked to tear the picture?" "How did you think the owner of the picture was going to feel when they saw it?" "How do you think another child would feel if they had torn the picture?" These questions are similar to those used in hypothetical vignettes, and might be helpful to clarify discrepancies between a child's actual behavior and their self-stated moral feelings and principles. It would also help to determine whether some children were experiencing empathy towards the victim but did not verbalize it due to fear of being chastised or a belief that they were supposed to do what they were told without question.

From an analytic perspective, our results also support the utility of a person-centered strategy in examining observational data of this nature. We did find interesting and significant associations between our single variables and composites derived from factor analysis and constructs of interest. However, there are a number of features of the Picture Tearing design that suggest the superiority of a person centered approach. First, demarcating behaviors into factors in this case seems somewhat artificial. Many of the variables demonstrating bodily tension and discomfort (e.g. Squirming, Hunched Shoulders) and apologizing to the victim did not load strongly onto any factor, yet these behaviors are theoretically important indicators of guilt and

internal aversive arousal (Kochanska & Aksan, 2006). When cluster analysis was utilized, we observed meaningful differences between bodily discomfort variables across clusters, which were consistent in magnitude and direction with empathy and guilt observed in each cluster, such that the cluster displaying the most guilt also displayed more bodily discomfort. This was a difference lost in factor analysis.

Furthermore, cluster analysis allowed us to examine coherent patterns of behavior and to speculate at the *meaning* of the behavior for that group. Factor analysis is predicated on the co-occurrence of variables, but is agnostic as to the potential that the co-occurrence of these variables within one person might be attributable to different factors than the co-occurrence of these variables within another person, and that these variables may actually demonstrate quite different levels of covariance when other factors are considered. Cluster analysis allows us to examine the possibility that the co-occurrence of the same variables in two different groups might actually have different correlates (and, thus, functional significance) depending on other characteristics of the members of those groups. Similarly, we can examine the possibility that variables that tend to co-occur in general across the sample actually do not co-occur within certain clusters, again due to other characteristics of the individuals in those clusters.

For instance, from a variable centered approach, we found that empathy was moderately, positively correlated with overall guilt and relief for the copy of the photo. However, members of the cluster with the highest mean for empathy actually demonstrated the lowest mean for guilt and relief. Moreover, despite expressing more empathy and defiance than High Compliers (the most compliant cluster), Low Compliers were actually shyer and less assertive, but somewhat poorer in their ability to shift attention. The factor analysis would lead us to believe that empathy is uniformly desirable, as it yields the adaptive response of guilt, whereas the cluster

analysis suggests that expressed empathy can actually reflect a subgroup of defiant, less inhibited children. Viewing the data in this manner allows us to consider more complex interrelationships among behaviors than is feasible from the perspective of factor analysis. It is suggested, then, that future research examining the Picture Tearing task utilize a cluster approach in lieu of or in addition to a factor analytic approach.

Limitations and future applications. One important limitation of the current study is our use of a community sample. Our sample is diverse in gender, ethnicity, and income level, which allows us to be reasonably confident that we have captured a range of behavior available in typically developing children. However, despite varying widely on demographic characteristics, our sample does not offer the range of behaviors that might be observed within a high-risk or clinical sample. Given that extremes in externalizing behavior and CU traits in the population are relatively low base rate (Canino, Polanczyk, Bauermeister, Rohde, & Frick, 2010), it is beyond the scope of a community sample to draw conclusions about children who score at extremes for EXT, CU traits, fearlessness, or other markers of severe disturbances in moral development. Indeed, the rates of externalizing pathology within our sample for all varieties of EXT are positively skewed.

Within this somewhat restricted range of EXT, we found that higher levels of verbalized empathy reflect oppositionality. The use of the Picture Tearing coding scheme in additional samples could provide further validation of the scheme as a useful measure of variation in child behaviors, particularly with respect to its utility for describing variations in EXT. As it stands, the codes have little relationship with EXT apart from empathic statements, which is associated with oppositionality. Given the robust link between low empathy and EXT in the literature, it is important to further investigate how the Picture Tearing behaviors relate to EXT in samples

where there is more variability in EXT. Within a higher risk sample, where rates of CU traits, EXT, and fearlessness are more extreme, it is possible that expressed empathy would instead separate those children with adequate perspective taking skills from those children with severely abnormal perspective taking. We might also expect to see higher rates of defiant and oppositional behavior across participants during the Picture Tearing task itself. Future research should examine these hypotheses in a high risk community or clinical sample.

However, there are also challenges to applying this paradigm to a high risk sample. In a higher risk sample, we would, as mentioned, expect to observe higher levels of externalizing pathology, including aggression and oppositionality, as well as lower levels of effortful control. Given the associations between empathic statements and oppositionality in our sample, it is possible that we would observe even stronger associations between these constructs in a high risk sample, given that these children should be even less likely to inhibit their responses and more likely to be defiant. Thus, it might be difficult to attribute empathic statements, and subsequence noncompliance, to actual concern for the victim, when these behaviors could just as easily reflect oppositionality on the part of the children. One important addition might be to more stringently test the nature of empathic statements in this group by examining the associations between empathic statements and purer measures of empathy and conscience, particularly the Inventory of Callous-Unemotional Traits (Kimonis, Frick, et al., 2008). If empathic statements in this sample are indeed indicative of concern for the victim and not purely oppositionality, then we would expect to observe a negative association between empathic statements and ICU scores.

Implications for the study of moral development. This study sought to validate a coding scheme designed to examine the facets of moral decision making in preschool and young school-age children, with a specific interest in connecting facets of moral behavior to

externalizing pathology. Unlike most recent work in the area of moral development and empathy, we utilized a laboratory task, which allowed us to observe a child's actual behavior in a naturalistic morally-relevant situation, rather than their responses to hypothetical situations. Using a scheme to numerically code the observed behaviors, we were able to examine contentions in the literature of what children say that they would do versus what they actually do given the opportunity. We have found that our coding scheme provides useful information about children's moral decision making that confirms, clarifies, and contests long-standing findings in the previous questionnaire-based literature. Thus, we were able to show, for example, that while older children are less likely to say that they would comply with a request from an authority figure if the request is perceived as being harmful (Damon, 1977; Laupa, 1991, 1994), in practice they are more likely to comply with such a request. Furthermore, sex differences in compliance and expressed perspective taking often found when analyzing responses to hypothetical vignettes (e.g. Garaigordobil, 2009; Hoffman, 1977) were largely absent in our observations. Given that previous research finding significant sex differences has been based on parent report, it is possible that parents in these studies are more likely to view girls as empathic and compliant in accordance with gender-based stereotypes. With our observational coding scheme, however, expression of empathy is rated objectively as number of statements made, which is putatively less likely to suffer gender-norm-based bias than subjective reporting of whether a child "seems" empathic or not to the rater.

Most importantly, our results provide additional support for the longstanding contention that the social press to comply with an authority figure can and often does supersede an individual's internalized moral rules to avoid harming another person. Additionally, these findings support work by Aksan and Kochanska (2005) and Kochanska and Aksan (2004)

suggesting that conscience is a higher order construct with two underlying components, moral emotions (empathy and guilt) and moral behavior. These components are correlated, but distinct, and each can occur to the exclusion of the other. Indeed, in our sample, as with Milgram's (1963, 1965) seminal studies and Shanab & Yahya's (1977) replication, the majority of children in our study complied with the experimenter's request to harm the victim's cherished property. Moreover, many of the children were compliant despite expressing concern for the victim's feelings or recognition that the request was immoral or illegitimate. In fact, older children, who putatively possess a better developed understanding of appropriate behavior and the victim's perspective, complied with the least delay and to the greatest extent.

On the whole, our data support the notion that the role of authority in the moral development of young children is complex, and that questionnaire-based measures may not adequately address this complexity on their own. Piaget (1965) and Kohlberg (Kohlberg & Kramer, 1969) seem to be accurate in their contention that an authority figure, whether present or internalized, is crucial to early moral development. The decision to act morally (or not) is contingent on the perceived consequences of the action. In our study, this notion manifests not only in the fact that most children complied with the authority despite any reservations they expressed (presumably due to fear of some imagined punishment), but also that their statements of reservation ostensibly reflect moral strictures gleaned from the adult authorities who have participated in their moral socialization. However, contrary to the assumptions of early theorists, the consequences that children of this age envision appear to be not only consequences to themselves in the form of punishment, but consequences to their victim.

Rather than passively receiving and executing adult commands, it appears that children view authority and their obligation to obey in nuanced ways (Kuczynski, Kochanska, Radke-

Yarrow, & Girnius-Brown, 1987). It does not appear to be a simple matter that "I must obey because the adult said so, and because I will be punished if I disobey." Rather, we observed that many children in our study seemed to evaluate the legitimacy of the authority figure critically. This is consistent with a small body of work that has found that children, when presented with hypothetical scenarios in which a parent or teacher commanded them to commit an immoral act (e.g. fight with another child), stated that the command was "wrong" and that they would not obey (Damon, 1977; Laupa, 1991, 1994). We did not find this to be the case. Rather, consistent with the only other in vivo study to utilize harmful commands (Shanab & Yahya, 1977), we found that most children did comply. However, we did observe that children, despite complying, seemed to evaluate critically the legitimacy of the authority figure's command, with the result that the degree and nature of their compliance varied. The more children expressed reservations about the "rightness" of the task, the longer they resisted the press to comply and the less thoroughly they carried out the command.

The way in which children interacted with and evaluated the authority figure as part of their decision making process appeared to interact in complex ways with age, in concert with increasing sophistication of moral reasoning and empathic perspective taking abilities, but also increasing efficacy of effortful control capabilities and awareness of social norms. Consistent with this notion, older children (five to seven years old) were more likely to openly question and disagree with the experimenter's agenda than younger children (three to four years old), and to express to the victim that they complied because they were commanded to do it. However, paradoxically, they were also more likely than younger children to comply quickly and completely. We had hypothesized that older children, with superior empathic skills, would be less likely to comply with the picture tearing request, despite being more likely to comply with

adult requests in general. This proved not to be the case. Our results suggest that, even though older children were more openly critical of the legitimacy of the command, they still tended to accept that the experimenter's authority to command them trumped their own preferences. Perhaps this explains why older children did not differ from younger children in their appearance of guilt; they might have believed that they were not personally responsible for the action, since they had acted against their will.

While younger children were less compliant with the request than older children, this did not appear to be a function of empathy—at least not observable empathy. Recall that younger children articulated fewer empathic statements than older children. One could presume that younger children experienced empathy internally, and that despite being more hesitant to express this empathy to the experimenter, it motivated them to defy the command. Yet, this explanation contradicts the theoretical notion that younger children do not have as sophisticated empathic abilities as older children and are more likely to base their moral decisions on the strictures of authority rather than abstract more principles than older children. However, empirically, multiple studies have found that younger children are less compliant with experimenter commands, potentially because they have poorer effortful control and are thus less able to modulate their behavior according to task expectations (Braungart-Rieker, Garwood, & Stifter, 1997; Kochanska & Aksan, 1995; Kochanska, Aksan, & Koenig, 1995). Thus, it appears that even though younger children are theoretically more likely to perceive authority figures as legitimate, their ability to comply with these authority figures is largely impacted other developmental limitations (including, potentially, aspects of effortful control).

Additionally surprising was the role that empathy and guilt played in predicting other behaviors of interest. Within the task itself, the children who demonstrated the most guilt were

likely to blame the other person and, unexpectedly, less likely to blame themselves. Guilt was also unrelated to apologizing or attempting to make reparations. It appears, then, that the children who experienced the most aversive arousal in the presence of their victim were not necessarily accordingly motivated to repair their relationship with the victim. It is possible that children experiencing high levels of guilt/anxious arousal were too focused on their own negative feelings to adequately focus on the feelings of the victim and what they might do as recompense. Instead, they seemed to attempt to modulate their aversive arousal by placing the blame on the authority figure who commanded them, or else withdrawing from the victim and remaining silent.

These putative processes are consistent with the notion that children who experience too much anxious arousal after a transgression might express low levels of guilt and empathy because they are too focused on their personal distress to adequately respond to the emotions of their victim (Eisenberg et al., 2010). This hypothesis in the literature is typically predicated on the assumption that excessive personal distress is a result of extreme dispositional fearfulness, high levels of anxiety, excessive power assertion on the part of the socialization agent (typically operationalized as a punitive parent), or a combination thereof (Kochanska, 1991). We did not find a link between fearfulness and empathic expression in our study. However, it is likely that the unique qualities of the picture tearing paradigm (having an authority figure ask the child to do something clearly harmful to another person, confronting the child with their victim) are strong enough situational presses to elicit excessive personal distress even in children who are not dispositionally fearful or anxious.

Not only did guilt not motivate reparative behavior during the task, but guilt and empathy were associated with externalizing pathology rated outside of the task—in particular, with

oppositionality. Moreover, these constructs were associated with deficits in areas of effortful control. These findings are unusual, given the wealth of research linking empathy and guilt to appropriate moral socialization, and thus low levels of externalizing pathology (Kochanska, 1991, 1993), as well as adequate levels of effortful control. Instead, it seems that children in our study who verbalized empathy were actually more oppositional and less inhibited than their peers. It is unlikely that empathy as defined in the literature is predictive of externalizing behaviors.

Rather, given the low-risk nature of this sample, it is probable that most children in our sample feel some degree of affective empathy towards the victim, and are able to take the victim's perspective. However, it appears that less impulsive, less oppositional children were able to override an empathic desire to refuse the experimenter in favor of the less intuitive behavior (ripping the photo) necessitated by task demands. On the contrary, children with more oppositional, impulsive temperaments were more likely to *express* that empathy in defiance of the experimenter's request, possibly because they were less able to inhibit the dominant response of refusing to commit an immoral act. In addition, children with more impulsive temperaments were possibly less likely to think through the consequences of their actions, and thus more likely to feel badly—or to fail to modulate their observable emotional response—after confronting the victim.

On the whole, our findings confirm that many modern notions of moral development in young children can be observed not only via parent report and children's assessment of hypothetical scenarios, but also in a naturalistic, in vivo scenario in which children believe they are facing an actual moral dilemma. In fact, using our coding scheme with the observational Picture Tearing task appears to provide additional information about children's moral behavior

than that which can be gleaned from questionnaire. It also confirmed many of the contentions previously hypothesized by theoretical and questionnaire-based research. We confirmed, for example, that empathy and guilt can be observed in children as young as three years of age, but that the presence of these moral emotions does not necessarily correlate strongly, or at all, with a child's enacted behavior. Indeed, the press to respond to an authority figure appears, as in previous observational studies, to trump a child's instinctive inclination to act prosocially. However, nor is the trade-off between empathy and compliance simple. Despite generally complying with the authority's immoral request, many children still evaluated the appropriateness of the task and resisted or modulated their level of compliance in ways that suggest a level of moral sophistication of which early theorists assumed them incapable.

Unexpectedly, we observed that children who expressed more empathy and guilt were also rated as showing more externalizing behaviors by their parents. This would seem to contradict the literature's contended associations between empathy, guilt, and adequate moral socialization, as operationalized by few externalizing problems. However, notably, the associations between empathy and externalizing pathology became weaker as children aged, suggesting that empathy in the Picture Tearing task is more reflective of concurrent oppositional behavior rather than a marker for enduring patterns of externalizing pathology. On the whole, in this age group and within the conditions of this task, compliance appears to be more adaptive than defiance, even though compliance is synonymous with committing an immoral or harmful act. Children who complied more extensively tended to be older, to express more empathy, and to demonstrate better effortful control. On the whole, our results suggest that moral decision making in preschool and school-aged children is a complex phenomenon affected by complex interrelationships between age, temperament, situational demands, and the child's own

internalized moral rules. Utilizing our coding scheme for the Picture Tearing task allowed us to observe subtleties in children's actual moral decision making that are not easily or not at all accessible via questionnaire.

APPENDICES

APPENDIX A: FIGURES

Figure 1. Diagram of moral development and its relationship to child internal and external characteristics and secondary outcomes.





Figure 2. Chart of Video Progression and Relevant Codes



Figure 3. Hierarchical cluster analysis agglomeration coefficients plotted against stage of analysis.



Figure 4. Gap statistic for hierarchical cluster analysis graphed against number of clusters.

Figure 5. Differences between cluster means for three cluster solutions.

Epoch 1.











Laughing

















Figure 5 (cont'd) Epoch 2.





Hunched Shoulders



Hiding





Smiling



Figure 5 (cont'd) Epoch 3.





Apologize












Figure 5 (cont'd)





Laughing



Figure 5 (cont'd)









APPENDIX B: TABLES

Table 1. In	traclass corre	lation coefficient	cients for v	video	coding	variables
-------------	----------------	--------------------	--------------	-------	--------	-----------

Epoch 1		Epoch 2		Epoch 3	
Variable	ICC	Variable	ICC	Variable	ICC
Empathy	0.89	Empathy	*	Blame Other	0.78
Hesitation	0.95	Hesitation	*	Blame Self	*
Defiance	0.98	Defiance	*	Apologize	0.85
Squirming	0.67	Squirming	*	Silence	0.67
Hunched Shoulders	0.23	Hunched Shoulders	0.23	Lying	0.7
Lip Biting	*	Lip Biting	*	Squirming	0.38
Crossed Arms	*	Crossed Arms	*	Hunched Shoulders	0.68
Hiding	*	Hiding	*	Lip Biting	0.01
Social Reference	0.57	Laughter	0.44	Crossed Arms	*
Laughter	0.93	Smiling	0.78	Hiding	0.99
Smiling	0.82			Social Reference	0.66
Number of Prompts	0.94			Laughter	0.84
How Much Torn	0.99			Smiling	0.91
How Much Torn Enjoyment	0.99 0.86			Smiling Gaze Avoidance	0.91 0.91
How Much Torn Enjoyment	0.99			Smiling Gaze Avoidance How Guilty	0.91 0.91 0.92

Note: * Insufficient variance among codes to calculate ICC

1. Request phase	2. After tear, before victim returns	3. After victim returns
Latency to tear	Squirming	Squirming
Tear/No tear	Lip biting	Lip biting
Number of prompts given	Hunched shoulders	Hunched shoulders
Amount of picture torn	Leaning away/Hiding	Leaning away/Hiding
Defiance/Noncompliance statements	Social referencing	Social referencing
Concern/empathy statements	Laughing / Smiling	Laughing / Smiling
Hesitation/questioning statements	Negative affect	Negative affect
Squirming	Crossing arms	Silence
Lip biting		Blaming other
Hunched shoulders		Lying
Leaning away/Hiding		Gaze avoidance
Social referencing		Global guilt rating
Laughing / Smiling		Blaming self

 Table 2. Picture Tearing codes relevant to each epoch of the task.

Negative affect	Apologizing
Enjoyment in tearing (reversed)	Relief at second picture
Crossing arms	Crossing arms

	U	
	Relevant analysis (if	
Number	applicable)	Hypothesis
<u>Aim 1</u>		
1.1		The task behaviors will be best represented by a three factor structure (see Table 4 for
1.1	EFA / CFA	specific hypothesized loadings)
1.2	CFA / Cluster	Most children will comply with the experimenter's request
1.3	CFA / Cluster	Noncompliance will be positively associated with Assertiveness (parent-rated)
1.4	CFA / Cluster	Noncompliance will be positively associated with expressed Empathy in the task
1.5		Noncompliance will be positively associated with Oppositionality (parent-rated ODD
1.5	CFA / Cluster	symptoms, Noncompliance in other lab tasks)
1.6		Noncompliance in the absence of empathic statements will be negatively associated with
1.0	CFA / Cluster	Fearlessness
Aim 2		
2.1	Correlations /	Noncompliance in PT will positively predict EXT, AGG, CU, Rule-breaking, ODD, and CD
2.1	Regressions	
2.2	Correlations /	Anxious Arousal in PT will negatively predict EXT, AGG, CU, Rule-breaking, ODD, and
2.2	Regressions	CD

 Table 3. Comprehensive list of hypotheses

2.2	Correlations /	Empathy in PT will negatively predict EXT, AGG, CU, Rule-breaking, ODD, and CD
2.5	Regressions	
2.4	Correlations /	Noncompliance will be modestly positively associated with child age
2.4	Regressions	
2.5	Correlations /	Anxious Arousal will be positively associated with child age
2.3	Regressions	
26	Correlations /	Empathy will be positively associated with child age
2.0	Regressions	
2.7	Correlations /	Noncompliance will be modestly negatively associated with female sex
	Regressions	
28	Correlations /	Anxious Arousal will be modestly positively associated with female sex
2.0	Regressions	
2.0	Correlations /	Empathy will be modestly positively associated with female sex
2.9	Regressions	
2 10	Correlations /	Anxious Arousal will be positively associated with total baseline Emotion Recognition score
2.10	Regressions	(more emotions correctly identified)
2.11	Correlations /	Empathy will be positively associated with total baseline Emotion Recognition score (more
2.11	Regressions	emotions correctly identified)

2.12	Correlations /	Anxious Arousal will be positively associated with PPVT score (better verbal ability)
2.12	Regressions	
2.12	Correlations /	Empathy will be positively associated with PPVT score (better verbal ability)
2.15	Regressions	
2.14	Correlations /	Noncompliance will be positively associated with Fearlessness
2.14	Regressions	
2.15	Correlations /	Anxious Arousal will be negatively associated with Fearlessness
2.13	Regressions	
2.16	Correlations /	Empathy will be negatively associated with Fearlessness
	Regressions	
		Empathy will have a curvilinear association with Fearlessness such that extreme high and
2.17	Correlations /	low Fearlessness will be associated negatively with Empathy, while moderate levels will be
	Regressions	associated positively with Empathy
2.19	Correlations /	Noncompliance will be positively associated with Anger, High Intensity Pleasure,
2.10	Regressions	
2.10	Correlations /	Anxious Arousal will be negatively associated with Anger, High Intensity Pleasure,
2.19	Regressions	
2.20	Correlations /	Empathy will be negatively associated with Anger, High Intensity Pleasure,
2.20	Regressions	

2.21	Correlations /	Noncompliance will be negatively associated with Sadness
2.21	Regressions	
2.22	Correlations /	Anxious Arousal will be positively associated with Sadness
2.22	Regressions	
2.22	Correlations /	Empathy will be positively associated with Sadness
2.23	Regressions	
2.24	Correlations /	Noncompliance will be negatively associated with Fear-proneness in other lab tasks
2.24	Regressions	
2.25	Correlations /	Anxious Arousal will be positively associated with Fear-proneness in other lab tasks
	Regressions	
2.26	Correlations /	Empathy will be positively associated with Fear-proneness in other lab tasks
2.20	Regressions	
2 27	Correlations /	Noncompliance will be negatively related to Effortful Control (high Inhibitory Control, low
2.27	Regressions	Impulsivity)
2.78	Correlations /	Anxious Arousal will be positively related to Effortful Control (high Inhibitory Control, low
2.20	Regressions	Impulsivity)
2 20	Correlations /	Empathy will be positively related to Effortful Control (high Inhibitory Control, low
2.27	Regressions	Impulsivity)

2.20	Correlations /	Noncompliance will be positively associated with Poor Monitoring, Corporal Punishment,
2.30 Regressions		Inconsistent Punishment
2.21	Correlations /	Anxious Arousal will be negatively associated with Poor Monitoring, Corporal Punishment,
2.31	Regressions	Inconsistent Punishment
Correlations /		Empathy will be negatively associated with Poor Monitoring, Corporal Punishment,
2.32	Regressions	Inconsistent Punishment
2.22	Correlations /	Noncompliance will be negatively related to Positive Parenting and Involvement
2.33	Regressions	
2.24	Correlations /	Anxious Arousal will be positively related to Positive Parenting and Involvement
2.34	Regressions	
2.25	Correlations /	Empathy will be positively related to Positive Parenting and Involvement
2.33	Regressions	
2.26	Correlations /	Within the PT task, empathy, defiance, hesitation, and signs of physical discomfort will be
2.30	Regressions	positively associated with Noncompliance and Guilt
<u>Aim 3</u>		
3.1	Growth Curve	Noncompliance in PT will positively and prospectively predict EXT, AGG, CU, Rule-
5.1	Analysis	breaking, ODD, and CD
3.2	Growth Curve	Anxious Arousal in PT will negatively and prospectively predict EXT, AGG, CU, Rule-
5.2	Analysis	breaking, ODD, and CD

	/	
2.2	Growth Curve	Empathy in PT will negatively and prospectively predict EXT, AGG, CU, Rule-breaking,
5.5	Analysis	ODD, and CD

 Table 4. Hypothesized results of factor analysis

Obedience/Compliance	Arousal	Empathy
Latency to tear	Hesitation statements	Concern/empathy statements
Defiance/Noncompliance statements	Squirming	Enjoyment in tearing (reversed)
Crossing arms	Lip biting	Relief at second picture
Tear/No tear	Hunched shoulders	
Number of prompts given	Leaning away/Hiding	
Amount of picture torn	Social referencing	
	Laughing / Smiling	
	Negative affect	
	Silence	
	Blaming other	
	Lying	
	Gaze avoidance	
	Global guilt rating	
	Blaming self	
	Apologizing	
	Hesitation/questioning	
	statements	

<u>Construct</u>	<u>Positive</u>	<u>Negative</u>
Compliance	Compliance in other tasks, Fearfulness, Effortful control, Inhibitory control	Assertiveness, High expressed empathy, Oppositionality, Age, Impulsivity, Harsh and inconsistent parenting
Empathy/ Anxious Arousal	Emotion recognition, Fearfulness, Sadness, Effortful control, PPVT	Guilt, Compliance, Fearlessness, EXT, Anger, Harsh and inconsistent parenting

Table 5. Hypothesized associations for variables investigated at baseline.

		Ep	och1				Epoch2			Ε	Epoch 3	
			Skewnes	<u>Kurtosi</u>	Mea	(SD		Kurtosi	Mea			Kurtosi
	Mean	<u>(SD)</u>	<u>s</u>	<u>s</u>	<u>n</u>)	Skewness	<u>s</u>	<u>n</u>	<u>(SD)</u>	Skewness	<u>s</u>
Empathy	1.15	(1.91)	2.23	5.16	0.02	0.13	7.49	55.50				
Hesitation	1.35	(1.68)	2.77	13.85	0.01	0.12	16.25	264.00				
Defiance	1.27	(2.47)	2.51	6.98	0.00	0.06	16.25	264.00				
Squirming	0.35	(1.11)	6.08	50.90	0.07	0.36	6.32	42.76	0.59	1.09	2.27	4.83
Hunched	0.22	(0.59)	3.01	20.49	0.18	0.67	5.10	31.53	0.46	0.95	2.75	8.43
Lin Biting	0.22	(0.37)	6.91	56.94	0.13	0.14	8 4 8	73 75	0.09	0.49	8.05	75 54
Crossing	0.07	(0.55)	0.71	50.74	0.12	0.14	0.40	13.13	0.07	0.47	0.05	75.54
Arms	0.09	(0.40)	5.53	32.23	0.03	0.27	12.56	176.71	0.07	0.25	3.92	14.76
Hiding	0.27	(0.80)	4.15	19.81	0.23	0.74	4.86	31.69	0.28	0.77	4.17	20.48
Social												
Referencin						0.00			0.12	0.31	3.30	12.84
g	1.56	(1.42)	1.89	6.17	0.00							
Laughing	0.26	(0.81)	4.37	21.75	0.02	0.14	6.43	40.22	0.24	0.66	3.76	16.81
Smiling	1.90	(2.38)	1.71	2.84	0.20	0.57	3.74	17.18	2.20	2.55	2.09	5.81
Tear or Not	0.78	(0.41)	-1.39	-0.08								
Number of												
Prompts	4.07	(2.68)	1.06	1.10								
Latency to		(43.48										
Tear	41.45)	1.29	2.18								
How Much												
Torn	3.36	(1.47)	-0.62	-1.11								
Enjoyment	2.03	(0.86)	0.52	-0.29								
Blame Other									0.95	1.09	1.14	0.96
Blame Self									0.18	0.45	2.78	8.15

 Table 6. Descriptive statistics for picture tearing variables.

Apologize					0.09	0.32	3.96	17.45
Silence					0.39	0.95	3.33	12.89
Lying					0.22	0.57	3.99	22.57
Gaze					0.41	1 14	1 81	31.06
Avoidance					0.71	1.17	т.0т	51.00
Overall					2 44	0.83	0.48	-0.54
Guilt					2.77	0.05	0.40	-0.54
Relief for					2.07	0.86	0.58	0.27
Сору					2.07	0.80	0.38	-0.27

		Epoch 2	2			Epoch 3	3					
_	Mean	<u>(SD)</u>	Skewness	<u>Kurtosis</u>	Mean	(SD)	Skewness	<u>Kurtosis</u>	Mean	(SD)	Skewness	<u>Kurtosis</u>
Empathy	0.83	1.05	0.91	-0.56	0.02	0.14	7.10	48.76				
Total Empathy	0.43	0.55	0.94	-0.46								
Hesitation	1.23	1.12	0.40	-1.20	0.01	0.12	16.25	264.00				
Total Hesitation	0.63	0.57	0.57	-0.33								
Defiance	0.72	1.05	1.10	-0.31	0.00	0.06	16.25	264.00				
Total Defiance	0.37	0.56	1.38	1.37								
Squirming	0.29	0.64	1.99	2.40	0.07	0.32	4.83	23.67	0.70	1.01	1.05	-0.38
Hunched Shoulders	0.27	0.57	2.02	2.96	0.16	0.48	3.11	8.55	0.63	1.05	1.38	0.34
Social Reference	1.61	1.17	0.48	-0.57					0.25	0.56	2.21	3.70
Total Social Reference	0.93	0.70	0.60	-0.22								
Laughing	0.15	0.36	1.94	1.77	0.03	0.16	5.93	33.39	0.43	0.99	2.09	2.62
Total Laughing	0.20	0.40	2.03	3.05								
Smiling	1.63	1.52	0.44	-1.27	0.19	0.49	2.56	5.73	2.02	1.73	0.44	-1.07
Total Smiling	1.30	1.06	0.55	-0.68								
Tear or Not	0.78	0.41	-1.39	-0.08								
Number of Prompts	3.60	1.74	0.00	-1.39								
Latency to Tear	41.45	43.48	1.29	2.18								
How Much Torn	3.30	1.46	-0.54	-1.17								
Enjoyment	2.18	0.88	0.30	-0.65								
Blame Other									0.95	1.03	0.68	-0.79

 Table 7. Descriptive statistics for recoded picture tearing variables.

Blame Self					0.21	0.48	2.26	4.45
Apologize					0.10	0.30	2.72	5.43
Silence					0.33	0.66	1.76	1.59
Lying					0.37	0.82	2.08	3.03
Gaze Avoidance					0.30	0.63	1.92	2.28
Overall Guilt					2.48	0.96	0.06	-0.93
Relief for Copy					1.98	0.88	0.68	-0.19

		1	2	3	4	5	6	7	8	9	10	11
1	Empathy	1.00										
2	Hesitation	0.39**	1.00									
3	Defiance	0.25^{**}	0.05	1.00								
4	Social Reference	0.07	0.12^{*}	-0.03	1.00							
5	Laughing	-0.08	-0.02	-0.13*	0.16**	1.00						
6	Smiling	-0.04	0.06	-0.10	0.22^{**}	0.46**	1.00					
7	Hunched Shoulders	0.02	0.02	0.03	0.14^{*}	0.06	0.19**	1.00				
8	Squirming	-0.04	-0.05	-0.03	0.24^{**}	0.12	0.29^{**}	0.14^{*}	1.00			
9	Tear or Not	-0.22**	-0.04	-0.60**	0.19**	0.18**	0.28^{**}	0.05	0.09	1.00		
10	Number of Prompts	0.32**	0.35**	0.63**	-0.01	-0.16*	-0.19**	-0.02	-0.15*	-0.56**	1.00	
11	Latency to Tear	0.18**	0.15^{*}	0.56**	0.00	-0.16**	-0.19**	-0.02	-0.15*	-0.58**	0.66**	1.00
12	How Much Torn	-0.27**	-0.08	-0.60**	0.08	0.17^{**}	0.24**	0.04	0.10	0.83**	-0.62**	-0.59**
13	Enjoyment	-0.13	-0.03	-0.13	0.11	0.39**	0.61**	0.12	0.14^{*}	0.16*	-0.18**	-0.15*
14	Blame Other	0.23**	0.20^{**}	-0.14*	0.21**	0.04	0.18^{**}	0.00	0.13*	0.35**	-0.06	-0.20**
15	Blame Self	-0.10	0.01	-0.10	-0.08	0.11	0.11	-0.13*	-0.03	0.21**	-0.01	-0.05
16	Apologize	0.14^{*}	0.15^{*}	-0.04	0.12	0.00	0.07	0.07	0.03	0.14^{*}	0.03	-0.03
17	Silence	-0.21**	-0.26**	-0.19**	0.07	-0.03	-0.10	0.07	0.14^{*}	0.26**	-0.34**	-0.16**
18	Lying	-0.14*	-0.03	-0.18**	0.04	0.11	0.09	0.12	-0.01	0.22^{**}	-0.22**	-0.15*
19	Gaze Avoidance	-0.03	-0.05	-0.09	0.06	-0.11	-0.14*	0.14^{*}	0.11	0.25**	-0.11	-0.12
20	Overall Guilt	0.26**	0.04	0.11	0.11	-0.05	-0.02	0.17^{*}	0.16*	0.06	0.02	0.00
21	Relief for Copy	0.20**	0.10	-0.03	0.06	0.07	0.34**	0.07	0.06	-0.01	-0.12	-0.14*

_

 Table 8. Bivariate correlations among picture tearing variables.

		12	13	14	15	16	17	18	19	20	21	
12	How Much Torn	1.00										
13	Enjoyment	0.15^{*}	1.00									
14	Blame Other	0.24**	0.05	1.00								
15	Blame Self	0.13*	0.12	-0.10	1.00							
16	Apologize	0.01	-0.03	0.13*	0.02	1.00						
17	Silence	0.27^{**}	-0.16*	-0.24**	-0.08	0.01	1.00					
18	Lying	0.28^{**}	0.06	-0.22**	0.00	-0.03	-0.01	1.00				
19	Gaze Avoidance	0.17**	-0.12	-0.03	-0.07	0.03	0.37**	0.06	1.00			
20	Overall Guilt	-0.05	-0.03	0.13	-0.16*	0.00	0.08	-0.05	0.24**	1.00		
21	Relief for Copy	-0.02	0.25**	0.16^{*}	-0.04	0.09	-0.22**	-0.03	-0.23**	0.30**	1.00	

		Demograph	nics		Emotion T	raining V	ariables				
					Total	Happy	Angry	Sad	Surprise	Fear	Neutral
		Age	<u>Sex</u>	<u>PPVT</u>	<u>Correct</u>	<u>Wrong</u>	<u>Wrong</u>	<u>Wrong</u>	<u>Wrong</u>	<u>Wrong</u>	<u>Wrong</u>
Epoch 1	Empathy	0.14*	-0.07	0.17^{**}	0.14*	-0.15*	-0.05	-0.13*	-0.10	-0.06	-0.13
	Hesitation	0.13*	0.02	0.11	0.08	-0.18**	-0.12	-0.09	-0.01	0.01	-0.04
	Defiance	0.00	-0.02	0.10	0.01	0.02	-0.03	0.03	0.02	-0.02	-0.05
	Squirming	0.02	0.07	0.00	-0.03	0.07	-0.03	0.01	0.00	0.03	0.04
	Hunched Shoulders	0.11	0.02	0.03	-0.07	0.12	0.06	0.00	0.08	0.04	0.01
	Social Ref	0.20^{**}	0.08	0.07	0.17^{**}	-0.19**	-0.07	-0.11	-0.22**	-0.11	-0.07
	Laugh	0.22^{**}	-0.11	0.07	0.16^{*}	-0.11	-0.12	-0.11	-0.06	-0.13*	-0.17**
	Smile	0.13	-0.06	-0.14	-0.07	0.04	-0.03	0.08	0.09	0.09	0.00
	Tear or Not	0.15	-0.07	-0.06	0.03	0.06	-0.01	0.06	-0.01	-0.03	-0.16*
	Latency	-0.11	-0.03	0.06	-0.04	0.03	0.04	0.01	0.03	0.06	0.02
	Number Prompts	-0.12*	-0.06	-0.03	-0.05	-0.01	0.03	0.00	0.09	0.06	0.03
	How Much Torn	0.15*	0.06	-0.03	0.08	-0.01	-0.04	-0.01	-0.07	-0.07	-0.11
	Enjoyment	0.12	0.00	-0.10	0.05	-0.06	-0.06	0.07	-0.04	-0.05	-0.07
Epoch 2	Empathy	-0.06	0.14^*	0.01	0.04	-0.04	-0.04	-0.05	0.00	0.00	-0.07
	Defiance	-0.03	0.06		0.04	-0.02	-0.02	-0.02	-0.03	-0.03	-0.03
	Squirming	0.14^{*}	0.06	-0.01	0.06	-0.05	-0.05	-0.06	-0.04	-0.03	-0.02
	Hunched Shoulders	0.14*	-0.09	0.01	0.02	-0.07	0.00	-0.03	0.00	0.01	-0.02
	Laugh	-0.05	-0.17***	0.00	0.00	0.07	-0.04	-0.06	0.05	-0.01	0.00

 Table 9. Correlations between picture tearing variables and child characteristics.

	Smile	0.14^{*}	0.01	0.08	0.05	-0.10	-0.10	0.00	0.00	0.02	-0.06
Epoch 3	Blame Other	0.23**	-0.06	0.05	0.11	-0.14*	-0.05	0.01	-0.06	-0.08	-0.16*
	Blame Self	-0.13*	-0.07	-0.12	-0.18**	0.00	0.09	0.20^{**}	0.13*	0.22^{**}	0.09
	Apologize	0.05	-0.05	-0.03	0.03	0.04	-0.03	-0.06	-0.06	-0.01	0.01
	Silence	0.02	0.10	0.02	0.02	0.11	-0.06	-0.07	-0.01	-0.04	0.00
	Lying	0.08	-0.03	0.02	-0.02	0.05	-0.01	0.02	0.02	0.01	0.01
	Squirming	0.08	0.19**	0.02	0.05	-0.02	-0.04	0.03	-0.04	-0.07	-0.07
	Hunched Shoulders	0.20^{**}	0.05	0.02	0.04	0.05	-0.05	0.00	-0.02	-0.06	-0.05
	Social Ref	0.13*	-0.06	0.01	0.08	-0.08	-0.09	-0.06	-0.05	-0.05	-0.05
	Laugh	0.09	0.10	0.02	0.03	-0.11	-0.02	0.05	-0.10	0.04	0.01
	Smile	0.31**	0.11	0.04	0.12	-0.09	-0.10	-0.02	-0.12	-0.08	-0.10
	Gaze Avoidance	0.03	-0.03	-0.02	-0.06	0.06	0.07	0.02	0.04	0.06	0.02
	Overall Guilt	0.12	0.10	0.05	0.05	0.02	0.00	-0.05	-0.10	-0.08	0.03
	Relief for Copy	0.10	0.19**	-0.03	0.02	-0.01	-0.02	0.01	-0.03	-0.01	-0.02
Totals	Empathy	0.11	-0.06	0.16^{*}	0.13*	-0.15*	-0.03	-0.14*	-0.09	-0.04	-0.12
	Hesitation	0.11	0.02	0.10	0.08	-0.18**	-0.11	-0.09	-0.01	0.01	-0.03
	Defiance	0.00	-0.02	0.10	0.00	0.01	-0.01	0.02	0.03	-0.01	-0.03
	Social Ref	0.17^{**}	0.01	0.04	0.11	-0.19**	-0.03	-0.10	-0.12	-0.04	-0.05
	Laugh	0.14^{*}	0.02	0.05	0.08	-0.12	-0.07	0.00	-0.10	-0.01	-0.05
	Smile	0.28^{**}	0.07	0.04	0.11	-0.08	-0.10	-0.01	-0.10	-0.06	-0.12
Composites	Guilt	0.13*	0.05	0.14^{*}	0.10	-0.10	-0.02	-0.12	-0.10	-0.05	-0.06
	Authority	0.20^{**}	-0.03	0.09	0.11	-0.20**	-0.09	-0.06	-0.04	-0.04	-0.12
	Noncompliance	-0.04	-0.02	0.05	-0.01	0.01	0.01	0.01	0.03	0.02	-0.05
	Positive Affect	0.23**	0.04	0.01	0.10	-0.11	-0.10	0.01	-0.10	-0.04	-0.11
	Avoidance	0.08	0.07	0.02	0.01	0.09	0.00	-0.04	-0.03	-0.02	0.01

		Mothe	r-report					Father-	report				
		<u>Pos</u> Parent	Involve	Consist	<u>Poor</u> Monitor	<u>Corp</u> Punish	<u>Approp</u> Punish	<u>Pos</u> Parent	Involve	Consist	<u>Poor</u> Monitor	<u>Corp</u> Punish	<u>Approp</u> Punish
Epoch 1	Empathy	0.12	0.14*	0.05	0.13	0.01	0.08	-0.02	-0.01	0.15	0.19*	0.18*	0.01
-	Hesitation	-0.02	0.05	0.02	0.06	0.02	0.02	0.01	0.03	0.01	0.08	0.03	-0.02
	Defiance	-0.04	-0.02	-0.12	0.14^{*}	-0.06	0.04	-0.02	-0.06	0.09	0.00	-0.08	0.05
	Squirming	-0.07	-0.12	0.05	0.01	-0.12	-0.03	0.05	-0.02	-0.02	-0.01	-0.08	0.01
	Hunched Shoulders	-0.08	-0.09	-0.01	0.07	-0.03	-0.19**	-0.12	-0.16	0.13	0.04	-0.09	-0.17
	Social Ref	0.00	0.08	0.03	0.20^{**}	-0.08	0.00	0.10	0.08	0.09	0.01	-0.07	-0.04
	Laugh	0.01	0.03	0.01	-0.10	0.04	0.13	0.03	0.04	0.05	-0.01	-0.07	0.05
	Smile	-0.11	-0.08	-0.07	0.07	0.08	-0.12	-0.07	0.00	-0.04	-0.05	-0.06	-0.10
	Tear or Not	-0.04	-0.03	0.08	0.00	0.03	-0.03	0.06	0.19	-0.19	0.13	0.13	0.04
	Latency	-0.06	-0.03	-0.08	0.11	-0.01	-0.03	-0.09	-0.16	0.12	-0.07	-0.11	0.00
	Number Prompts	-0.03	-0.03	-0.12	0.13	-0.05	0.04	-0.12	-0.13	0.02	-0.02	-0.08	0.07
	How Much Torn	-0.07	-0.04	0.10	0.02	0.03	-0.05	0.02	0.09	-0.11	0.04	0.11	-0.04
	Enjoyment	-0.08	-0.10	-0.02	0.00	0.13	0.01	0.11	0.08	0.18	-0.02	-0.05	0.06
Epoch 2	Empathy	0.05	0.03	-0.13	-0.06	0.00	0.03	0.09	0.06	-0.05	-0.01	0.03	0.03
	Defiance	0.02	0.02	0.01	0.02	0.00	-0.01	. c	. ^c	°.		.c	. c
	Squirming	-0.01	-0.01	0.03	0.05	0.11	0.03	-0.03	0.03	0.03	-0.02	0.13	0.12

 Table 10. Correlations between picture tearing variables and parenting characteristics.

	Hunched Shoulders	0.03	0.04	0.05	0.25**	0.01	0.05	-0.10	-0.12	0.12	-0.05	-0.05	0.06
	Laugh	-0.17*	-0.21**	0.03	-0.03	0.00	-0.11	0.03	-0.01	0.06	-0.06	-0.05	0.03
	Smile	-0.03	-0.01	0.04	0.03	0.02	0.02	-0.07	-0.01	-0.05	0.04	-0.10	0.00
Epoch 3	Blame Other	-0.02	0.01	0.04	0.06	-0.05	0.06	0.11	0.12	0.05	0.12	0.03	0.04
	Blame Self	0.01	-0.04	-0.06	-0.08	0.02	0.02	-0.10	-0.09	0.02	-0.14	-0.06	-0.06
	Apologize	0.00	0.04	0.10	0.06	-0.08	-0.03	0.06	0.12	0.02	0.13	0.16	0.09
	Silence	-0.05	0.00	0.06	0.05	0.00	178*	-0.04	-0.01	-0.11	-0.02	0.09	-0.14
	Lying	0.00	0.00	-0.01	-0.10	0.03	0.02	-0.07	-0.02	0.08	0.04	0.05	-0.14
	Squirming	-0.05	0.00	0.00	-0.11	-0.05	-0.05	0.10	0.13	-0.10	0.03	-0.01	-0.06
	Hunched Shoulders	0.01	0.10	0.01	0.18^{*}	0.07	0.01	-0.03	-0.01	-0.08	-0.03	-0.04	-0.07
	Social Ref	-0.10	-0.06	0.05	0.08	0.14	0.03	0.00	0.04	0.00	0.13	0.05	0.01
	Laugh	0.01	0.00	0.01	-0.13	-0.03	0.06	-0.06	-0.04	0.09	0.07	0.03	-0.14
	Smile	-0.13	-0.02	0.10	-0.01	0.04	0.03	-0.03	-0.03	0.07	0.08	-0.01	-0.18*
	Gaze Avoidance	0.01	0.02	0.07	0.05	-0.07	165*	-0.11	-0.06	-0.19*	0.04	0.04	-0.01
	Overall Guilt	0.10	0.11	-0.09	0.16^{*}	-0.08	-0.09	0.14	0.06	0.03	0.09	0.10	-0.13
	Relief for Copy	0.03	0.04	0.04	0.08	-0.05	-0.13	0.02	0.01	0.17	0.05	-0.01	-0.10
Totals	Empathy	0.11	0.13	0.03	0.12	0.02	0.08	-0.03	-0.03	0.16	0.18^{*}	0.18^{*}	0.01
	Hesitation	-0.03	0.05	0.02	0.07	0.01	0.03	0.01	0.02	0.02	0.08	0.03	-0.03
	Defiance	-0.06	-0.03	-0.13	0.13*	-0.08	0.03	-0.05	-0.10	0.12	0.00	-0.07	0.02
	Social Ref	-0.06	0.02	0.03	0.15^{*}	0.02	0.04	0.04	0.04	0.15	0.07	-0.04	-0.03
	Laugh	-0.02	-0.02	0.01	-0.14*	-0.03	0.07	-0.05	-0.03	0.11	0.05	-0.01	-0.11
	Smile	-0.12	-0.03	0.05	0.03	0.05	0.01	-0.04	-0.02	0.05	0.08	0.00	-0.11
Composites	Guilt	0.10	0.15*	-0.02	0.19**	-0.05	-0.04	0.04	0.02	0.15	0.15	0.13	-0.10

Authority	-0.02	0.05	0.04	0.08	-0.03	0.06	0.08	0.09	0.04	0.12	0.04	0.00
Noncompliance	-0.11	-0.06	-0.11	0.19**	-0.06	0.00	-0.12	-0.14	0.08	-0.02	-0.08	0.03
Positive Affect	-0.10	-0.06	0.02	-0.04	0.04	0.02	-0.02	-0.02	0.13	0.07	-0.01	-0.09
Avoidance	0.00	0.05	0.05	0.11	-0.04	-0.19**	-0.01	0.01	-0.14	0.06	0.12	-0.11

	Three Factor	Four Factor	Five Factor	Rational Structure
Empathy	Guilt	Guilt	Guilt	Empathy
Hesitation	Guilt	Guilt	Authority	Arousal
Defiance	Noncompliance	Noncompliance	Noncompliance	Compliance
Social Reference	(dropped)	(dropped)	(dropped)	Arousal
Laughing	Positive Affect	Positive Affect	Positive Affect	Arousal
Smiling	Positive Affect	Positive Affect	Positive Affect	Arousal
Hunched				
Shoulders	(dropped)	(dropped)	(dropped)	Arousal
Squirming	(dropped)	(dropped)	(dropped)	Arousal
Number of				
Prompts	Noncompliance	Noncompliance	Noncompliance	Compliance
Latency to Tear	Noncompliance	Noncompliance	Noncompliance	Compliance
How Much Torn	Noncompliance	Noncompliance	Noncompliance	Compliance
Enjoyment	Positive Affect	Positive Affect	Positive Affect	Empathy
Blame Other	Guilt	Guilt	Authority	Arousal
Blame Self	(dropped)	(dropped)	(dropped)	Arousal
Apologize	(dropped)	(dropped)	(dropped)	Arousal
Silence	Noncompliance	Avoidance	Avoidance	Arousal
Lying	(dropped)	(dropped)	(dropped)	Arousal
Gaze Avoidance	Noncompliance	Avoidance	Avoidance	Arousal
Overall Guilt	Guilt	Guilt	Avoidance/Guilt	Arousal
Poliof for Conv		Guilt/Positive		
Relief for Copy	Guilt	Affect	Guilt	Empathy

Table 11. Factor structures based on exploratory factor analysis.

Model	Description	x2 (df)	prob	CFI	TLI	RMSEA	SRMR
1	3 factor	310.27 (74)	<.00	0.77	0.72	0.11	0.11
2	4 factor	251.97 (71)	<.00	0.82	0.77	0.1	0.1
3	5 factor	216.87 (65)	<.00	0.85	0.79	0.09	0.09
4	4 factor after modifications	207.23 (69)	<.00	0.86	0.82	0.09	0.09
5	5 factor after modifications	200.53 (65)	<.00	0.87	0.81	0.09	0.09
6	Pationally derived 3 factor	No					
0	Kationally delived 5 factor	convergence					

 Table 12. Fit statistics and comparative analyses for confirmatory factor analyses.

Factor	Variable	Estimate	S.E.	p-value
Authority	Hesitation	1.00	0.00	-
	Blame Other	0.70	0.28	0.01
	How Many			
	Prompts	1.17	0.29	< 0.01
Noncompliance	Defiance	1.00	0.00	-
	Latency to tear	76.46	6.35	< 0.01
	How Many			
	Prompts	3.22	0.24	< 0.01
	How Much Torn	-2.60	0.21	< 0.01
Positive Affect	Laughter	1.00	0.00	-
	Smiling	4.10	0.55	< 0.01
	Enjoyment	3.05	0.41	< 0.01
Avoidance	Silence	1.00	0.00	-
	Gaze Avoidance	0.63	0.16	< 0.01
	How Guilty	1.19	1.10	0.03
Guilt	Empathy	1.00	0.00	-
	How Guilty	2.39	2.26	0.03
	Relief for Picture	1.27	0.74	0.07

 Table 13. Factor loading estimates for final five factor model

						Hunched	Social			Latency	Tearor	Number	How Much	
E	ooch 1	Empathy	Hesitation	Defiance	Squirming	Shoulders	Reference	Laughing	Smiling	to Tear	Not	Prompts	Torn	Enjoyment
1	Mean	1.04	1.51	1.16	0.40	0.30	1.21	0.12	0.76	50.83	0.66	4.72	2.84	1.40
	Std. Dev	1.16	1.17	1.11	0.74	0.58	1.04	0.33	1.02	17.19	0.48	1.20	1.61	1.19
2	Mean	1.23	1.60	1.84	0.47	0.47	1.19	0.12	0.35	124.00	0.30	5.56	1.60	0.63
	Std. Dev	1.25	1.26	1.07	0.77	0.77	1.10	0.32	0.69	31.64	0.46	0.63	1.07	0.95
3	Mean	0.62	1.00	0.22	0.19	0.20	1.31	0.18	0.92	15.16	0.96	2.54	4.01	2.14
	Std. Dev	0.89	1.00	0.61	0.53	0.49	0.95	0.38	1.03	8.44	0.19	1.29	1.04	1.01
E	boch 2	Empathy	Hesitation	Defiance	Squirming	Hunched Shoulders	Social Reference	Laughing	Smiling					
1	Mean	0.01	0.00	0.00	0.01	0.01	0.04	0.00	0.09					
	Std. Dev	0.12	0.00	0.00	0.12	0.12	0.21	0.00	0.29					
2	Mean	0.00	0.00	0.00	0.02	0.14	0.02	0.02	0.07					
	Std. Dev	0.00	0.00	0.00	0.15	0.35	0.15	0.15	0.26					
3	Mean	0.03	0.00	0.01	0.04	0.05	0.09	0.04	0.13					

Table 14. Descriptive statistics for three cluster solution.

	Std. Dev	0.16	0.00	0.08	0.21	0.22	0.28	0.19	0.33					
E	ooch 3	Blame Other	Blame Self	Apologize	Silence	Lying	Squirming	Hunched Shoulders	Social Reference	Laughing	Smiling	Gaze Avoid	Overall Guilt	Relief for Copy
1	Mean	0.88	0.22	0.07	0.22	0.28	0.55	0.43	0.22	0.18	1.55	0.24	1.64	1.37
	Std. Dev	1.02	0.49	0.26	0.57	0.65	0.93	0.86	0.55	0.67	1.61	0.58	1.41	1.20
2	Mean	0.53	0.19	0.09	0.12	0.09	0.09	0.30	0.19	0.21	1.05	0.12	0.72	0.58
	Std. Dev	0.88	0.55	0.29	0.45	0.43	0.37	0.80	0.50	0.77	1.25	0.45	1.22	1.01
3	Mean	1.08	0.20	0.11	0.43	0.47	0.91	0.79	0.26	0.58	2.42	0.36	2.36	2.14
	Std. Dev	1.04	0.45	0.31	0.72	0.93	1.08	1.14	0.58	1.12	1.77	0.67	1.05	0.98

				,					,	1				
	Agg	RB	ODD	CD	INT	EXT	CU	Agg	RB	ODD	CD	INT	EXT	CU
Guilt Composite	0.15*	0.11	0.17^{*}	0.08	0.06	0.14^{*}	0.08	0.16	0.07	0.11	0.10	-0.05	0.13	0.08
Authority Composite	0.08	0.09	0.10	0.07	0.07	0.09	0.08	0.05	-0.06	0.03	-0.05	-0.03	0.01	0.06
Noncompliance Composite	-0.02	0.01	0.07	-0.04	0.05	-0.01	0.00	0.07	0.00	0.10	0.03	0.09	0.05	0.03
Pos Affect Composite	0.06	0.07	0.00	0.09	-0.06	0.07	0.10	0.03	0.05	0.00	0.00	-0.13	0.04	0.10
Avoidance Composite	-0.03	0.00	-0.09	0.01	-0.03	-0.02	0.11	-0.08	-0.06	-0.17	-0.02	-0.16	-0.08	0.03
Total Empathy	0.17^{*}	0.09	0.23**	0.06	0.11	0.15^{*}	0.05	0.17	0.09	0.15	0.12	-0.02	0.15	0.13
Total Hesitation	0.07	0.13	0.10	0.09	0.04	0.10	0.12	0.04	-0.01	0.03	-0.02	-0.04	0.02	0.04
Total Defiance	-0.01	0.04	0.06	0.00	0.09	0.01	-0.05	0.09	0.10	0.14	0.11	0.14	0.10	-0.02
Total Social Reference	0.14^{*}	0.16^{*}	0.07	0.16*	0.12	0.16*	0.09	0.01	-0.07	-0.04	0.01	-0.10	-0.02	0.03
Total Laughing	0.08	0.03	0.07	0.05	-0.06	0.06	0.07	0.04	0.08	0.01	0.02	-0.10	0.06	0.11
Total Smiling	0.02	0.07	-0.04	0.07	-0.03	0.04	0.12	-0.05	-0.01	-0.07	-0.05	-0.10	-0.03	0.00
Latency to Tear	-0.08	-0.02	0.00	-0.09	0.01	-0.06	-0.02	-0.02	-0.05	0.04	-0.03	0.08	-0.03	-0.04
Tear or Not	-0.01	-0.13	-0.08	-0.06	-0.02	-0.07	0.08	-0.10	-0.17	-0.14	-0.15	-0.16	-0.13	-0.06
How Much Torn	0.04	-0.01	-0.04	0.04	0.00	0.02	0.09	-0.05	-0.07	-0.10	-0.06	-0.17	-0.06	0.06
Enjoyment in Tearing	0.07	0.14	0.00	0.16*	-0.08	0.11	0.07	0.09	0.10	0.09	0.08	-0.10	0.10	0.08
Gaze Avoidance	-0.04	-0.04	-0.09	0.00	-0.05	-0.04	0.08	0.02	-0.04	-0.08	0.06	-0.09	0.00	0.03
Overall Guilt	0.10	0.11	0.07	0.09	-0.03	0.11	0.10	0.02	0.03	-0.04	0.01	-0.10	0.02	0.09
Relief for Copy	0.08	0.07	0.08	0.07	-0.01	0.08	0.03	0.11	0.14	0.06	0.12	-0.01	0.13	0.01
Child variables														

 Table 15. Correlations between task variables, child and environmental characteristics, and problem behaviors.

Child Age	0.09	0.16^{*}	0.02	0.17^{*}	0.26**	0.13	0.09	-0.06	0.03	-0.08	-0.01	0.14	-0.03	-0.03
Child Sex	-0.09	-0.07	-0.02	-0.13	0.08	-0.09	0.00	-0.12	-0.13	-0.11	-0.16	0.10	-0.13	-0.04
PPVT Score	-0.06	0.01	-0.02	-0.01	0.16^{*}	-0.04	-0.08	-0.06	0.01	-0.03	-0.03	0.12	-0.04	-0.07
Mother-report														
Positive Parenting	-0.16*	0.21**	-0.15*	0.21**	-0.09	- 0.19 ^{**}	-0.17*	-0.18*	-0.18*	-0.18*	-0.15	-0.12	-0.19*	0.37**
Involvement	- 0.19 ^{**}	0.20**	- 0.18 ^{**}	0.21**	-0.02	0.20**	-0.14*	0.25**	0.24**	0.25**	0.23**	-0.06	0.27**	0.34**
Inconsistent Discipline	0.25*	0.23**	0.24**	0.20**	0.13	0.26**	0.27**	0.04	0.16	0.03	0.14	-0.05	0.09	0.10
Poor Monitoring	0.19**	0.14*	0.17^{*}	0.15*	0.17^{*}	0.19**	0.28^{**}	0.05	0.06	-0.01	0.08	0.06	0.06	0.06
Corporal Punishment	0.28**	0.22**	0.24**	0.23**	0.14^{*}	0.27**	0.11	0.15	0.24**	0.12	0.23**	0.07	0.20^{*}	0.09
Appropriate Punishment	0.19**	0.09	0.23**	0.07	0.13	0.16*	0.11	-0.11	-0.10	-0.04	-0.14	-0.21*	-0.11	-0.18*
Father-report														
Positive Parenting	0.23**	-0.05	- 0.30 ^{**}	-0.08	-0.13	-0.17	-0.12	0.27**	-0.20*	0.32**	-0.21*	-0.17*	0.26**	-0.19*
Involvement	-0.19*	-0.01	- 0.28 ^{**}	-0.07	-0.07	-0.12	-0.08	- 0.33 ^{**}	-0.20*	- 0.38 ^{**}	- 0.24 ^{**}	-0.15	- 0.30 ^{**}	0.22**
Inconsistent Discipline	0.16	0.11	0.22^{*}	0.06	0.09	0.15	-0.07	0.24**	0.24**	0.26**	0.21^{*}	0.19^{*}	0.26**	0.16
Poor Monitoring	0.03	0.05	0.08	-0.01	0.07	0.04	0.12	-0.05	-0.03	-0.02	-0.07	0.11	-0.05	0.15
Corporal Punishment	0.08	0.03	0.05	0.01	-0.04	0.06	0.17	0.10	0.10	0.04	0.06	-0.01	0.10	0.20^{*}

Appropriate	0.01	0.01	0.04	-0.01	0.01	0.01	0.09	-0.07	-0.06	-0.03	-0.10	-0.03	-0.07	-0.04
Funishinent														
Emotion Recognition														
# Emotions														
# Elliotions	0.06	0.06	-0.01	0.07	0.07	0.06	-0.02	-0.14	-0.03	-0.13	-0.08	-0.03	-0.11	-0.14
Hanny Wrong	-0.02	-0.01	0.00	-0.02	-0.02	-0.02	0.11	0.03	-0.04	0.03	-0.04	-0.02	0.00	0.03
Angry Wrong	-0.02	-0.01	0.00	-0.02	-0.02	-0.02	0.11	0.03	-0.04	0.03	-0.04	-0.02	0.00	0.03
	0.01	0.00	0.05	0.01	-0.03	0.01	0.02	0.20	0.05	0.18	0.13	0.08	0.16	0.29
Sadness Wrong	-0.05	-0.07	0.01	-0.07	-0.05	-0.06	0.02	0.06	-0.03	0.07	0.06	-0.01	0.03	-0.02
Surprised Wrong	-0.09	-0.09	-0.02	-0.09	-0.03	-0.10	-0.03	0.14	0.06	0.13	0.08	0.05	0.12	0.23*
Fear Wrong	-0.01	0.00	0.05	-0.03	-0.05	-0.01	0.06	0.11	0.06	0.10	0.08	0.01	0.10	0.10
Neutral Wrong	-0.08	-0.06	-0.04	-0.08	-0.10	-0.08	-0.05	0.06	0.02	0.06	0.04	0.00	0.05	0.02
Mother-report														
-														
Effortful	0.22*	-	-	-	0.00	-	-	-	0.01*	-	0.00*	0.04	-	-
Effortful Control	-0.33*	0.32**	- 0.32 ^{**}	- 0.35 ^{**}	-0.06	- 0.35 ^{**}	- 0.30 ^{**}	- 0.26 ^{**}	-0.21*	- 0.26 ^{**}	-0.22*	-0.04	0.26**	- 0.26 ^{**}
Effortful Control Negative Emotionality	-0.33* 0.21**	0.32**	0.32 ^{**} 0.24 ^{**}	0.35 ^{**} 0.09	-0.06 0.30**	0.35 ^{**} 0.17 [*]	0.30 ^{**}	0.26 ^{**} 0.31 ^{**}	-0.21* 0.11	0.26 ^{**} 0.30 ^{**}	-0.22* 0.18*	-0.04 0.32**	0.26 ^{**} 0.25 ^{**}	0.26 ^{**} 0.17 [*]
Effortful Control Negative Emotionality Positive Emotionality	-0.33* 0.21** 0.29**	0.32** 0.09 0.16*	0.32** 0.24** 0.28**	0.35** 0.09 0.17*	-0.06 0.30** -0.11	0.35** 0.17* 0.25**	0.30** 0.07 0.05	0.26** 0.31** 0.36**	-0.21* 0.11 0.27**	0.26** 0.30** 0.39**	-0.22* 0.18* 0.32**	-0.04 0.32** -0.06	0.26** 0.25** 0.35**	0.26 ^{**} 0.17 [*] 0.20 [*]
Effortful Control Negative Emotionality Positive Emotionality Father-report	-0.33* 0.21** 0.29**	0.32 ^{**} 0.09 0.16 [*]	0.32 ^{**} 0.24 ^{**} 0.28 ^{**}	0.35 ^{**} 0.09 0.17 [*]	-0.06 0.30** -0.11	0.35 ^{**} 0.17 [*] 0.25 ^{**}	0.30** 0.07 0.05	0.26** 0.31** 0.36**	-0.21* 0.11 0.27**	0.26** 0.30** 0.39**	-0.22* 0.18* 0.32**	-0.04 0.32** -0.06	0.26 ^{**} 0.25 ^{**} 0.35 ^{**}	0.26 ^{**} 0.17 [*] 0.20 [*]
Effortful Control Negative Emotionality Positive Emotionality <i>Father-report</i> Effortful	-0.33* 0.21** 0.29**	0.32** 0.09 0.16*	0.32 ^{**} 0.24 ^{**} 0.28 ^{**}	0.35 ^{**} 0.09 0.17 [*]	-0.06 0.30** -0.11	0.35 ^{**} 0.17 [*] 0.25 ^{**}	0.30** 0.07 0.05	0.26** 0.31** 0.36**	-0.21* 0.11 0.27**	0.26** 0.30** 0.39**	-0.22* 0.18* 0.32**	-0.04 0.32** -0.06	0.26** 0.25** 0.35**	0.26 ^{**} 0.17 [*] 0.20 [*]
Effortful Control Negative Emotionality Positive Emotionality <i>Father-report</i> Effortful Control	-0.33* 0.21** 0.29** 0.34**	0.32 ^{**} 0.09 0.16 [*] -0.20 [*]	0.32 ^{**} 0.24 ^{**} 0.28 ^{**} 0.33 ^{**}	0.35 ^{**} 0.09 0.17 [*] -0.22 [*]	-0.06 0.30** -0.11 -0.11	0.35 ^{**} 0.17 [*] 0.25 ^{**} 0.30 ^{**}	0.30** 0.07 0.05 -0.22*	0.26** 0.31** 0.36** 0.35**	-0.21* 0.11 0.27** 0.26**	0.26** 0.30** 0.39** 0.38**	-0.22* 0.18* 0.32** 0.29**	-0.04 0.32** -0.06 -0.11	0.26** 0.25** 0.35** 0.34 ^{**}	0.26 ^{**} 0.17 [*] 0.20 [*] 0.29 ^{**}
Effortful Control Negative Emotionality Positive Emotionality <i>Father-report</i> Effortful Control Negative Emotionality	-0.33* 0.21** 0.29** 0.34** 0.18*	0.32** 0.09 0.16* -0.20* 0.10	$\begin{array}{c} 0.32^{**} \\ 0.24^{**} \\ 0.28^{**} \\ \hline \\ 0.33^{**} \\ 0.21^{*} \end{array}$	0.35 ^{**} 0.09 0.17 [*] -0.22 [*] 0.15	-0.06 0.30** -0.11 -0.11 0.24**	0.35 ^{**} 0.17 [*] 0.25 ^{**} 0.30 ^{**} 0.16	0.30** 0.07 0.05 -0.22* 0.01	0.26^{**} 0.31^{**} 0.36^{**} 0.35^{**} 0.41^{**}	-0.21* 0.11 0.27** 0.26** 0.18*	0.26^{**} 0.30^{**} 0.39^{**} 0.38^{**} 0.40^{**}	-0.22* 0.18* 0.32** 0.29** 0.28**	-0.04 0.32** -0.06 -0.11 0.32**	0.26** 0.25** 0.35** 0.34** 0.34**	0.26** 0.17* 0.20* 0.20* 0.29** 0.25**
Effortful Control Negative Emotionality Positive Emotionality <i>Father-report</i> Effortful Control Negative Emotionality Positive Emotionality	-0.33* 0.21** 0.29** 0.34** 0.18*	0.32** 0.09 0.16* -0.20* 0.10 0.07	$\begin{array}{c} - \\ 0.32^{**} \\ 0.24^{**} \\ 0.28^{**} \\ \hline \\ 0.33^{**} \\ 0.21^{*} \\ 0.21^{*} \\ \end{array}$	0.35 ^{**} 0.09 0.17 [*] -0.22 [*] 0.15 0.14	-0.06 0.30** -0.11 -0.11 0.24** -0.08	0.35 ^{**} 0.17 [*] 0.25 ^{**} 0.30 ^{**} 0.16 0.15	0.30** 0.07 0.05 -0.22* 0.01 -0.09	$\begin{array}{c} - \\ 0.26^{**} \\ 0.31^{**} \\ 0.36^{**} \\ \hline \\ 0.35^{**} \\ 0.41^{**} \\ 0.32^{**} \end{array}$	-0.21* 0.11 0.27** 0.26** 0.18* 0.21*	$\begin{array}{c} - \\ 0.26^{**} \\ 0.30^{**} \\ 0.39^{**} \\ \hline \\ 0.38^{**} \\ 0.40^{**} \\ 0.31^{**} \end{array}$	-0.22* 0.18* 0.32** 0.29** 0.28** 0.25**	-0.04 0.32** -0.06 -0.11 0.32** -0.07	$\begin{array}{c} 0.26^{**} \\ 0.25^{**} \\ 0.35^{**} \\ \hline 0.34^{**} \\ 0.34^{**} \\ 0.30^{**} \end{array}$	$\begin{array}{c} - \\ 0.26^{**} \\ 0.17^{*} \\ 0.20^{*} \\ \hline \\ 0.29^{**} \\ 0.25^{**} \\ 0.26^{**} \end{array}$

Fearfulness	-0.08	-0.07	-0.03	-0.09	-0.04	-0.08	-0.05	0.00	-0.01	0.08	-0.01	0.07	0.00	-0.05
Compliance	-0.15	-0.13	-0.14	-0.14	0.05	-0.15	-0.15	0.29*	0.26*	0.31**	0.27**	-0.07	0.30*	-0.17

	Mother-r	eported	-										
	Anger	<u>Inhib</u> <u>Cont</u>	<u>Att</u> <u>Shift</u>	<u>Shy</u>	<u>Hi</u> <u>Pleas</u>	<u>Impul</u>	Soothe	Fear	<u>Att</u> Focus	<u>Sadness</u>	EC	<u>NE</u>	PE
Guilt Composite	0.04	-0.03	-0.08	-0.10	0.12	0.23**	-0.05	0.01	-0.02	0.06	0.01	0.07	0.18**
Authority Composite	0.10	-0.08	-0.16*	-0.13	0.14*	0.22^{**}	-0.10	0.08	-0.08	0.05	-0.09	0.09	0.14*
Noncomplianc e Composite	0.07	-0.09	-0.17*	0.19**	-0.13	-0.05	-0.04	0.11	- 0.18 ^{**}	0.08	-0.18*	0.09	0.01
Pos Affect Composite	-0.03	-0.04	0.10	-0.12	0.04	0.08	-0.04	-0.04	0.03	-0.13	0.01	-0.09	-0.02
Avoidance Composite	-0.09	0.06	0.17*	0.03	-0.02	-0.01	0.10	-0.08	0.10	-0.11	0.04	-0.08	-0.08
Total Empathy	0.08	-0.02	- 0.19 ^{**}	-0.10	0.05	0.18**	-0.05	0.00	-0.03	0.11	0.01	0.08	0.12
Total Hesitation	0.14^{*}	-0.13	-0.15*	0.00	-0.01	0.12	-0.09	0.13	-0.12	0.08	-0.11	0.14*	0.09
Total Defiance	-0.03	-0.04	-0.10	0.13	-0.10	-0.05	0.01	0.15^{*}	-0.12	0.04	-0.14*	0.06	-0.03
Total Social Reference	0.03	-0.13	-0.05	-0.05	0.13	0.13	-0.10	-0.02	-0.03	-0.08	-0.09	-0.03	0.08
Total Laughing	0.07	-0.02	-0.01	-0.06	0.06	0.04	-0.10	-0.04	0.00	0.01	0.02	-0.02	0.00
Total Smiling	-0.10	0.02	0.10	-0.08	-0.02	0.05	0.01	-0.04	0.07	-0.13	0.07	-0.12	-0.04
Latency to Tear	0.11	-0.10	0.21**	0.22^{**}	- 0.18 ^{**}	-0.07	-0.04	0.08	-0.17*	0.12	-0.13	0.10	0.02
Tear or Not	-0.07	0.07	0.07	-0.10	0.09	0.00	-0.02	-0.14	0.10	-0.17^{*}	0.07	-0.19*	-0.11
How Much Torn	-0.08	0.04	0.15*	-0.14*	0.17^{*}	0.05	-0.01	-0.14*	0.09	-0.15*	0.06	-0.16*	-0.02

 Table 16. Correlations between picture tearing variables and temperament traits.

Table	16	(cont'd)

Enjoyment in Tearing	-0.08	-0.12	0.17*	-0.19*	0.08	0.16*	0.03	-0.01	-0.07	-0.23**	-0.08	-0.11	0.02
Gaze Avoidance	-0.01	0.02	0.11	0.03	-0.08	-0.04	0.06	-0.01	0.03	-0.03	-0.03	0.01	-0.13
Overall Guilt	0.02	0.02	0.04	-0.01	0.10	0.16*	0.02	-0.01	-0.04	-0.03	-0.02	0.03	0.15
Relief for Copy	0.07	-0.07	0.06	-0.01	0.10	0.14	-0.12	0.04	-0.10	0.05	-0.01	0.09	0.17*
	Father-re	ported											
	<u>Anger</u>	<u>Inhib</u> <u>Cont</u>	<u>Att</u> <u>Shift</u>	<u>Shy</u>	<u>Hi</u> <u>Pleas</u>	<u>Impul</u>	Soothe	Fear	<u>Att</u> <u>Focus</u>	<u>Sadness</u>	<u>EC</u>	<u>NE</u>	<u>PE</u>
Guilt Composite	0.01	-0.16	-0.01	-0.12	0.16	0.21^{*}	-0.04	-0.04	-0.13	0.17	-0.16	-0.01	0.12
Authority Composite	-0.01	-0.02	-0.03	-0.13	0.03	0.12	-0.07	0.01	-0.04	-0.03	-0.10	-0.08	0.03
Noncomplianc e Composite	0.07	-0.05	-0.07	0.11	-0.15	-0.01	0.05	-0.03	-0.11	0.09	-0.11	0.09	-0.05
Pos Affect Composite	-0.05	-0.01	0.20^{*}	-0.08	0.12	0.03	0.00	-0.02	0.07	-0.09	0.03	-0.07	0.04
Avoidance Composite	-0.05	-0.01	0.09	-0.02	0.02	0.07	-0.01	0.10	0.08	-0.02	0.05	0.00	0.02
Total Empathy	0.02	-0.17	-0.15	-0.11	0.05	0.17	-0.05	-0.06	-0.14	0.14	-0.17*	-0.02	0.05
Total Hesitation	0.05	-0.07	0.00	-0.02	-0.03	0.08	0.01	-0.03	-0.08	0.00	-0.16	-0.03	0.02
Total Defiance	0.06	0.02	-0.05	0.05	-0.11	-0.05	0.09	-0.01	-0.06	0.01	-0.03	0.07	-0.06
Total Social Reference	0.06	-0.16	-0.04	0.04	0.17	0.13	-0.15	0.17*	-0.09	-0.03	-0.06	0.07	0.18*
Total Laughing	-0.01	-0.03	0.08	-0.07	0.11	-0.01	0.06	-0.06	0.00	-0.04	-0.02	-0.05	0.03
Total Smiling	-0.10	0.05	0.21^{*}	-0.03	0.02	-0.01	0.01	-0.05	0.11	-0.12	0.09	-0.11	-0.01

Latency to Tear	0.08	-0.02	-0.08	0.20^{*}	-0.18*	-0.08	0.04	-0.10	-0.12	0.17^{*}	-0.10	0.07	-0.05				
Tear or Not	-0.03	0.08	0.10	-0.02	0.16	0.12	0.08	0.09	0.25*	-0.11	0.18	-0.03	0.13				
How Much Torn	-0.10	-0.02	0.11	-0.15	$.020^{*}$	0.20^{*}	-0.05	0.07	0.06	-0.12	0.03	-0.08	0.12				
Enjoyment in Tearing	-0.08	-0.04	0.20^{*}	-0.11	0.15	0.13	0.00	0.05	-0.02	-0.01	-0.01	0.01	0.11				
Gaze Avoidance	0.08	-0.03	0.04	0.02	-0.12	-0.04	-0.01	0.20^{*}	0.10	0.04	0.05	0.13	-0.07				
Overall Guilt	-0.03	-0.03	0.08	-0.04	0.16	0.16	-0.03	0.04	-0.04	0.11	-0.06	0.03	0.16				
Relief for Copy	0.05	-0.17	0.19	-0.02	0.10	0.12	0.01	-0.11	-0.12	0.11	-0.10	-0.03	0.10				
			Effects	s of pred	ictors							Effects	s of pred	lictors			
---------	------------	------------	----------	-----------	----------	--------------	------------------	----------	----------	--------	------------	----------	-----------	----------	--------------	------------------	----------
							ΔR^2 for	block								ΔR^2 for	block
Mother	<u>r</u>						ΔF		Father							ΔF	
reporte	e <u>d</u>	_	<u>B</u>	<u>t</u>	<u>p</u>	ΔR^2		<u>p</u>	reported		_	<u>B</u>	<u>t</u>	<u>p</u>	ΔR^2		<u>p</u>
AGG									AGG								
	Step 1					.037	1.54	.178		Step 1					.023	0.58	.713
		Guilt	0.06	2.40	.017						Guilt	0.04	1.38	.172			
		Authority	0.00	0.15	.883						Authority	0.00	-0.10	.920			
		Noncomp	-0.02	-0.41	.683						Noncomp	0.00	-0.09	.926			
		Pos Affect	0.02	0.68	.495						Pos Affect	0.00	0.10	.921			
		Avoidance	-0.03	-0.92	.359						Avoidance	-0.04	-1.22	.227			
RB	Step 1					.027	5.70	.018	RB								
	I	Age	0.02	1.94	.054												
	Step 2					.012	0.49	.783		Step 1					.015	0.38	.865
		Guilt	0.02	1.22	.224						Guilt	0.02	0.91	.363			
		Authority	0.01	0.28	.780						Authority	-0.02	-0.83	.409			
		Noncomp	0.00	0.13	.894						Noncomp	-0.01	-0.42	.679			
		Pos Affect	0.01	0.58	.565						Pos Affect	0.01	0.31	.757			
		Avoidance	-0.01	-0.26	.798						Avoidance	-0.02	-0.97	.332			
ODD									ODD								
	Step 1					.055	2.33	.044		Step 1					.042	1.10	.366
		Guilt	0.10	2.75	.006						Guilt	0.06	1.19	.236			
		Authority	0.01	0.26	.794						Authority	0.00	0.02	.983			
		Noncomp	0.02	0.38	.707						Noncomp	0.01	0.12	.906			
		Pos Affect	0.00	-0.01	.993						Pos Affect	-0.02	-0.41	.683			
		Avoidance	-0.07	-1.68	.095						Avoidance	-0.12	-2.09	.038			

Table 17. Hierarchical multiple regression analyses predicting current child problem behaviors from picture tearing taskcomposite variables

		,															
CD	Step 1					.030	6.42	.012	CD								
		Age	0.02	2.53	.012												
	Step 2					.010	0.41	.845		Step 1					.010	0.26	.936
		Guilt	0.02	1.05	.295						Guilt	0.02	0.90	.368			
		Authority	0.00	0.04	.965						Authority	-0.02	-0.85	.399			
		Noncomp	-0.01	-0.33	.745						Noncomp	-0.01	-0.20	.840			
		Pos Affect	0.01	0.74	.457						Pos Affect	0.00	0.01	.993			
		Avoidance	-0.01	-0.26	.794						Avoidance	-0.01	-0.56	.578			
INT	Step 1					.067	14.68	.000	INT								
		Age	0.03	3.83	.000												
	Step 2					.025	1.08	.373		Step 1					.049	1.29	.273
		Guilt	0.01	0.59	.556						Guilt	-0.01	-0.33	.745			
		Authority	0.00	0.02	.987						Authority	0.00	0.02	.984			
		Noncomp	0.00	0.18	.861						Noncomp	0.01	0.20	.845			
		Pos Affect	-0.03	-1.91	.057						Pos Affect	-0.03	-1.56	.120			
		Avoidance	-0.02	-1.07	.288						Avoidance	-0.04	-1.71	.090			
EXT									EXT								
											Age						
	Step 1					.032	1.35	.245		Step 1					.020	0.50	.775
		Guilt	0.04	2.08	.039						Guilt	0.03	1.28	.203			
		Authority	0.01	0.38	.702						Authority	-0.01	-0.40	.691			
		Noncomp	-0.01	-0.21	.835						Noncomp	-0.01	-0.23	.820			
		Pos Affect	0.02	0.88	.381						Pos Affect	0.00	0.19	.849			
		Avoidance	-0.01	-0.58	.560						Avoidance	-0.03	-1.20	.233			
CU									CU								
	Step 1					.034	1.43	.216		Step 1					.011	0.28	.923
		Guilt	0.01	0.50	.618						Guilt	0.00	-0.10	.922			
		Authority	0.02	0.74	.462						Authority	0.03	0.90	.371			
		Noncomp	0.02	0.54	.587						Noncomp	0.00	-0.07	.947			

Pos Affect	0.04	1.71	.089			Pos Affect	0.01	0.40	.690		
Avoidance	0.04	1.67	.097			Avoidance	0.01	0.38	.705		

			Effects of pr	edictors			ΔR^2 for bl	ock
			B	<u>t</u>	p	ΔR^2	ΔF	p
AGG	Step 1					.039	4.20	.016
		Empathy	0.08	2.27	.024			
		Social Ref	0.05	1.58	.115			
RB	Step 1					.033	3.56	.030
		Age	0.02	1.30	.198			
	Step 2					.017	3.64	.058
		Social Ref	0.04	1.91	.058			
ODD	Step 1					.052	11.21	.001
		Empathy	0.17	3.35	.001			
CD	Step 1					.021	3.53	.062
		Enjoyment	0.03	1.88	.062			
EXT	Step 1					.034	3.71	.026
		Social Ref	0.04	1.79	.076			

 Table 18. Hierarchical multiple regression analyses predicting current child problem behaviors from single picture tearing task variables

Table 19. Hierarchical multiple regression analyses predicting current child temperament traits from picture tearing tas	k
composite variables	

			Effects predic	s of tors			ΔR^2 for block	or				Effects predict	s of tors			ΔR^2 for block	r
Lab task variables			B	<u>t</u>	р	ΔR^2	F	p		_		B	t	р	ΔR^2	F	p
Fear	Step 1					.01 2	1.92	.168	Comp	Step 1					.11 8	26.6 4	.000
		Age	-0.63	-1.38	.168						Age	0.15	5.16	.00. 0			
	Step 2					.02 4	0.77	.576		Step 2					.05 0	2.31	.046
		Guilt	1.23	1.58	.115						Guilt	0.04	0.78	.43 5			
		Authority	-1.09	-1.37	.172						Authority	-0.03	-0.68	.50 0			
		Noncomp	-0.01	-0.01	.994						Noncomp	-0.16	-1.98	.04 9			
		Pos Affect	0.25	0.29	.769						Pos Affect	-0.02	-0.32	.75 0			
		Avoidanc e	-0.61	-0.69	.489						Avoidanc e	0.11	1.81	.07 1			
Mother reported									Father reported								
Anger	Step 1					.03 6	3.80	.024	Anger	Step 1					.06 4		.015
		Age	-0.13	-2.60	.010						Age	-0.16	-2.58	.01 1			
	Step 2					.02 5	1.08	.374		Step 2					.01 3		.875
		Guilt	0.03	0.34	.731						Guilt	-0.01	-0.13	.89 7			
		Authority	0.17	1.78	.077						Authority	0.13	1.10	.27 3			
		Noncomp	0.03	0.23	.819						Noncomp	0.01	0.07	.94 6			

		Pos Affect	-0.02	-0.23	.815						Pos Affect	-0.05	-0.43	.66 9		
		Avoidanc e	-0.07	-0.76	.446						Avoidanc e	-0.07	-0.55	.58 6		
Inhib	Step 1					.05 7	6.16	.003	Inhib	Step 1					.09 5	.002
Contr		Age	0.13	3.13	.002				Contr		Age	0.15	2.75	.00 7		
	Step 2					.03 1	1.37	.236		Step 2					.04 1	.323
		Guilt	0.00	-0.06	.954						Guilt	-0.17	-1.80	.07 4		
		Authority	-0.12	-1.48	.139						Authority	-0.07	-0.66	.51 2		
		Noncomp	-0.14	-1.11	.270						Noncomp	-0.01	-0.08	.93 7		
		Pos Affect	-0.13	-1.65	.100						Pos Affect	-0.09	-0.85	.39 5		
		Avoidanc e	0.00	0.03	.976						Avoidanc e	0.05	0.42	.67 8		
Att Shift	Step 1								Att Shift	Step 1						
											Age					
	Step 2					.08 5		.003		Step 2					.05 0	.228
		Guilt	-0.09	-0.96	.339						Guilt	-0.05	-0.40	.69 1		
		Authority	-0.20	-1.97	.051						Authority	-0.14	-1.15	.25 3		
		Noncomp	-0.22	-1.51	.133						Noncomp	0.02	0.10	.92 2		
		Pos Affect	0.12	1.24	.215						Pos Affect	0.26	2.03	.04 4		
		Avoidanc e	0.20	1.96	.051						Avoidanc e	0.19	1.36	.17 6		
Shyness	Step 1								Shyness	Step 1						
											Age					

-	\rightarrow	<u></u>							1	1					
	Step 2					.07 9	.003		Step 2					.05 1	 .236
		Guilt	-0.18	-1.64	.102					Guilt	-0.18	-1.29	.19 9		
		Authority	-0.17	-1.38	.168					Authority	-0.19	-1.29	.20 0		
		Noncomp	0.49	2.71	.007					Noncomp	0.23	1.06	.29 1		
		Pos Affect	-0.18	-1.59	.114					Pos Affect	-0.12	-0.78	.43 9		
		Avoidanc e	0.09	0.76	.446					Avoidanc e	0.03	0.18	.86 0		
Hi Pleas	Step 1							Hi Pleas	Step 1					.10 2	.001
		Age								Age	0.10	1.62	.10 7		
	Step 2					.05 4	.035		Step 2					.05 1	.205
		Guilt	0.15	1.86	.064					Guilt	0.18	1.77	.07 9		
		Authority	0.12	1.35	.178					Authority	-0.04	-0.37	.70 9		
		Noncomp	-0.32	-2.40	.017					Noncomp	-0.28	-1.75	.08 2		
		Pos Affect	0.02	0.20	.844					Pos Affect	0.10	0.87	.38 7		
		Avoidanc e	-0.07	-0.77	.443					Avoidanc e	-0.07	-0.55	.58 7		
Impuls	Step 1							Impuls	Step 1						
		Age								Age					
	Step 2	0				.11 6	.000		Step 2	0				.07 8	.057
		Guilt	0.22	3.01	.003					Guilt	0.21	2.38	.01 9		
		Authority	0.20	2.58	.011					Authority	0.11	1.18	.24 1		

		Noncomp	-0.14	-1.21	.229					Noncomp	0.00	-0.01	.99 3		
		Pos Affect	0.13	1.74	.083					Pos Affect	0.07	0.72	.47 5		
		Avoidanc e	-0.01	-0.08	.940					Avoidanc e	0.03	0.27	.78 9		
Soothe	Step 1							Soothe	Step 1						
		Age								Age					
	Step 2					.02 0	.532		Step 2					.02 0	.774
		Guilt	-0.05	-0.80	.426					Guilt	-0.02	-0.22	.82 7		
		Authority	-0.04	-0.59	.555					Authority	-0.09	-1.03	.30 5		
		Noncomp	-0.04	-0.41	.686					Noncomp	0.16	1.21	.22 7		
		Pos Affect	-0.01	-0.18	.857					Pos Affect	0.03	0.37	.71 4		
		Avoidanc e	0.10	1.39	.165					Avoidanc e	0.02	0.15	.88 0		
Fear	Step 1							Fear	Step 1						
	1	Age								Age					
	Step 2					.02 4	.399		Step 2					.02 5	.679
		Guilt	-0.05	-0.53	.598					Guilt	-0.14	-1.18	.24 2		
		Authority	0.10	0.96	.336					Authority	0.02	0.19	.84 8		
		Noncomp	0.17	1.17	.245					Noncomp	-0.05	-0.26	.79 6		
		Pos Affect	-0.07	-0.78	.438					Pos Affect	-0.05	-0.38	.70 7		
		Avoidanc e	-0.08	-0.83	.407					Avoidanc e	0.19	1.37	.17 3		
Att Focus	Step 1					.05 3	.004	Att Focus	Step 1					.02 4	.211

		Age	0.10	2.47	.014					Age	0.03	0.52	.60 7		
	Step 2					.05 4	.039		Step 2					.03 9	.403
		Guilt	-0.02	-0.25	.803					Guilt	-0.14	-1.50	.13 7		
		Authority	-0.09	-1.06	.290					Authority	-0.01	-0.13	.89 3		
		Noncomp	-0.33	-2.71	.007					Noncomp	-0.13	-0.92	.36 1		
		Pos Affect	-0.09	-1.19	.234					Pos Affect	0.01	0.14	.88 8		
		Avoidanc e	0.04	0.53	.599					Avoidanc e	0.13	1.21	.22 9		
Sad	Step 1							Sad	Step 1						
		Age								Age					
	Step 2					.04 3	.107		Step 2					.04 5	.328
		Guilt	0.08	1.12	.263					Guilt	0.18	2.13	.03 5		
		Authority	0.04	0.46	.648					Authority	-0.07	-0.80	.42 3		
		Noncomp	0.02	0.19	.848					Noncomp	0.07	0.53	.59 6		
		Pos Affect	-0.16	-2.07	.039					Pos Affect	-0.04	-0.49	.62 5		
		Avoidanc e	-0.14	-1.70	.090					Avoidanc e	-0.07	-0.75	.45 3		
EC	Step 1							EC	Step 1						
		Age								Age					
	Step 2					.04	.103		Step 2					.05 4	.208
		Guilt	0.04	0.74	.463					Guilt	-0.10	-1.61	.11 0		
		Authority	-0.09	-1.51	.131					Authority	-0.08	-1.27	.20 6		

Table 1	9 (cont'	d)													
		Noncomp	-0.21	-2.49	.014					Noncomp	-0.06	-0.60	.55 2		
		Pos Affect	-0.05	-0.92	.356					Pos Affect	-0.01	-0.15	.88 0		
		Avoidanc e	-0.02	-0.41	.685					Avoidanc e	0.06	0.87	.38 7		
NE	Step 1							NE	Step 1						
		Age								Age					
	Step 2					.03 2	.236		Step 2					.00 8	.956
		Guilt	0.03	0.68	.495					Guilt	-0.02	-0.33	.74 2		
		Authority	0.07	1.36	.176					Authority	-0.02	-0.43	.67 0		
		Noncomp	0.03	0.48	.630					Noncomp	0.06	0.70	.48 4		
		Pos Affect	-0.06	-1.33	.186					Pos Affect	-0.02	-0.27	.78 6		
		Avoidanc e	-0.05	-0.95	.342					Avoidanc e	0.01	0.10	.91 8		
PE	Step 1							PE	Step 1						
		Age								Age					
	Step 2					.06 4	.017		Step 2					.01 7	.797
		Guilt	0.10	2.76	.006					Guilt	0.06	1.31	.19 4		
		Authority	0.05	1.42	.157					Authority	0.00	-0.02	.98 3		
		Noncomp	-0.04	-0.74	.462					Noncomp	-0.03	-0.47	.63 9		
		Pos Affect	0.00	0.07	.943					Pos Affect	0.03	0.51	.61 2		
		Avoidanc e	-0.05	-1.30	.196					Avoidanc e	-0.02	-0.29	.77 5		

Appendix C. Original picture tearing coding scheme.

ID: _____ Coder: _____ Date: _____

Durbin Lab Guilt Coding Scheme

Time - first prompt given: ______ Time - child rips the picture / experimenter gives up: _____

Code timestamps for verbalizations: (examples below)

Concern: It's his favorite, What if she's mad, She'll be sad / Hesitation: Are you sure? Can I? / Defiance: No, I can't, I'm not going to

Concern/empathy	Hesitation	Defiance/Noncompliance

Bodily reactions (AFTER PROMPT but BEFORE ALBUM OWNER RETURNS)

Squirming	Hunched shoulders	Lip biting	Crossing arms
Leaning	Social reference	Laughing	Smiling
away/Hiding			

Did the child tear the picture when prompted? How many prompts had to be given?

- No
- Yes
- How many prompts: _____

How much did the child tear the picture?

- 1. Not at all (refused to tear)
- 2. A small amount (the corner, a small rip)
- 3. A moderate amount (Torn in some places, around edges)
- 4. A large amount (Torn into several pieces, in half)
- 5. Most of the picture (torn into pieces, can't see the picture at all)

How much enjoyment did the child exhibit in tearing the picture?

- 1. None
- 2. A little
- 3. A moderate amount
- 4. A lot

AFTER ALBUM OWNER RETURNS

How the child responded when the experimenter came back?

Blame other	Blame self	Apologize	Silence (every five	Lying
			seconds)	

Bodily reactions

Squirming	Hunched shoulders	Lip biting	Crossing arms
Leaning	Social reference	Laughing	Smiling
away/Hiding			

Gaze avoidance (code every five seconds of averted gaze)

How guilty does the child seem?

- 1. Not at all
- 2. A little
- 3. Moderate
- 4. A lot

How much relief does the child show at the second picture?

- 1. None
- 2. A little
- 3. Moderate amount
- **4. A lot**

Other comments:

Appendix D. Description of clusters in four cluster solution.

and differences between clusters, we conducted a one-way ANOVA using cluster membership and the variables used to create the clusters as dependent variables. We also conducted post-hoc comparisons of significant ANOVAs using Tukey's HSD. Average characteristics of the four clusters and characteristics on which the clusters differed are described below.

Description of clusters in four cluster solution. In order to determine the similarities

Cluster 1. Members of this cluster comprised 24.9% of the sample (n = 67). Compared to other clusters, children in cluster 1 took more time to comply and tore the picture more completely than children in cluster 3, but took less time and tore less than children in clusters 2 and 4. During the first epoch (between the first prompt and the main experimenter leaving), they showed more empathy and hesitation than children in cluster 3 but did not differ from clusters 2 and 4 on these variables. They displayed less defiance than cluster 4 and more defiance than cluster 3. In response to the victim, they showed less physical discomfort (i.e. squirming, hunched shoulders) than cluster 3 and more than clusters 2 and 4. They also displayed more guilt and less relief than cluster 3, but less guilt and more relief than clusters 2 and 4.

Cluster 2. Members of this cluster comprised 9.3% of the sample (n = 25). Children in cluster 2 differed from children in cluster 4 only on Latency to tear (d = 2.98) and Relief in seeing the photo copy (d = 0.67). Compared to the other clusters, they showed more empathy, hesitation, and defiance than cluster 3 and tore the picture less completely than clusters 1 and 3. In response to the victim, they showed less physical discomfort (i.e. squirming, hunched shoulders) and less guilt than clusters 1 and 3. They were also less likely to lie or be silent than cluster 3.

Cluster 3. Members of this cluster comprised 59.1% of the sample (n = 159). Compared to other clusters, children in this cluster waited the shortest amount of time before tearing the

picture and were the most likely to tear the picture. During the first epoch, they displayed less empathy and hesitation than clusters 1 and 2 and less defiance than all the other clusters. They also tore the picture more completely than the other clusters and demonstrated the most enjoyment while tearing. Notably, they also displayed the most guilt upon the victim's return and the most relief on receiving the copy of the picture. Also in response to the victim, they were the most likely to be silent and to appear visibly uncomfortable (i.e. squirming, hunched shoulders).

Cluster 4. Members of this cluster comprised 6.7% of the sample (n = 18). Children in cluster 4 differed from children in cluster 2 only on Latency to tear (d = 2.98) and Relief in seeing the photo copy (d = 0.67). Compared to the other clusters, they showed more empathy, hesitation, and defiance than cluster 3 and tore the picture less completely than clusters 1 and 3. In response to the victim, they showed less physical discomfort (i.e. squirming, hunched shoulders) and less guilt than clusters 1 and 3. They were also less likely to lie or be silent than cluster 3.

Appendix E. Recommended final picture tearing coding scheme retaining most empirically useful variables.

ID: _____ Coder: _____ Date: _____

Durbin Lab Guilt Coding Scheme

Time - first prompt given: _____

Time - child rips the picture / experimenter gives up: _____

Code timestamps for verbalizations: (examples below)

Concern: It's his favorite, What if she's mad, She'll be sad / Hesitation: Are you sure? Can I? / Defiance: No, I can't, I'm not going to

Concern/empathy	Hesitation	Defiance/Noncompliance

Bodily reactions (AFTER PROMPT but BEFORE ALBUM OWNER RETURNS)

Social reference	Laughing	Smiling	

Did the child tear the picture when prompted? How many prompts had to be given?

- No
- Yes
- How many prompts: _____

How much did the child tear the picture?

- 6. Not at all (refused to tear)
- 7. A small amount (the corner, a small rip)
- 8. A moderate amount (Torn in some places, around edges)
- 9. A large amount (Torn into several pieces, in half)
- 10. Most of the picture (torn into pieces, can't see the picture at all)

How much enjoyment did the child exhibit in tearing the picture?

- 5. None
- 6. A little
- 7. A moderate amount
- 8. A lot

AFTER EXPERIMENTER LEAVES, BEFORE EXPERIMENTER COMES BACK When child is alone in room

Code timestamps for verbalizations: (examples below)

Concern: It's his favorite, What if she's mad, She'll be sad / Hesitation: Are you sure? Can I? / Defiance: No, I can't, I'm not going to

Concern/empathy	Hesitation	Defiance/Noncompliance

Bodily reactions (AFTER PROMPT but BEFORE ALBUM OWNER RETURNS)

Social reference	Laughing	Smiling

AFTER ALBUM OWNER RETURNS

How the child responded when the experimenter came back?

Blame other	Blame self	Apologize	Silence (every five	Lying
			seconds)	

Bodily reactions

Social reference	Laughing	Smiling	

Gaze avoidance (code every five seconds of averted gaze)

How guilty does the child seem?

- 5. Not at all
- 6. A little
- 7. Moderate
- 8. A lot

How much relief does the child show at the second picture?

- 5. None
- 6. A little
- 7. Moderate amount
- 8. A lot

Other comments:

REFERENCES

REFERENCES

- Abe, J., & Izard, C. (1999). Compliance, Noncompliance Strategies, and the Correlates of Compliance in 5-year-old Japanese and American Children. *Social Development*. Retrieved from http://onlinelibrary.wiley.com/doi/10.1111/1467-9507.00077/full
- Achenbach, T., & Rescorla, L. (2001). *ASEBA school-age forms & profiles*. Retrieved from http://aseba.com/ordering/ASEBA Reliability and Validity-School Age.pdf
- Aksan, N., & Kochanska, G. (2005). Conscience in Childhood: Old Questions, New Answers. *Developmental Psychology*, 41(3), 506–516. Retrieved from http://eric.ed.gov/?id=EJ684980
- Asendorpf, J. B., & Nunner-Winkler, G. (1992). Children's Moral Motive Strength and Temperamental Inhibition Reduce Their Immoral Behavior in Real Moral Conflicts. *Child Development*, 63(5), 1223–1235. http://doi.org/10.1111/j.1467-8624.1992.tb01691.x
- Belacchi, C., & Farina, E. (2012). Feeling and Thinking of Others: Affective and Cognitive Empathy and Emotion Comprehension in Prosocial/Hostile Preschoolers. *Aggressive Behavior*, 38(2), 150–165. http://doi.org/10.1002/ab.21415
- Beyers, J. M., Bates, J. E., Pettit, G. S., & Dodge, K. A. (2003). Neighborhood structure, parenting processes, and the development of youths' externalizing behaviors: A multilevel analysis. *American journal of community psychology*, *31*(1-2), 35-53.
- Blair, R. J., Colledge, E., Murray, L., & Mitchell, D. G. (2001). A selective impairment in the processing of sad and fearful expressions in children with psychopathic tendencies. *Journal of Abnormal Child Psychology*, 29, 491–498. http://doi.org/10.1023/A:1012225108281
- Blair, R. J. R., Budhani, S., Colledge, E., & Scott, S. (2005). Deafness to fear in boys with psychopathic tendencies. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, 46(3), 327–36. http://doi.org/10.1111/j.1469-7610.2004.00356.x
- Blair, R. J. R., & Coles, M. (2000). Expression recognition and behavioural problems in early adolescence. *Cognitive Development*, 15(4), 421–434. http://doi.org/10.1016/S0885-2014(01)00039-9
- Blass, T. (1991). Understanding behavior in the Milgram obedience experiment: The role of personality, situations, and their interactions. *Journal of Personality and Social Psychology*. Retrieved from http://psycnet.apa.org/journals/psp/60/3/398/
- Borke, H. (1971). Interpersonal Perception of Young Children: Egocentrism or Empathy?. *Developmental Psychology*. Retrieved from http://eric.ed.gov/?id=EJ043606

- Borke, H. (1973). The Development of Empathy in Chinese and American Children between Three and Six Years of Age: A Cross-Culture Study. *Developmental Psychology*. Retrieved from http://eric.ed.gov/?id=EJ081919
- Braungart-Rieker, J., Garwood, M., & Stifter, C. (1997). Compliance and noncompliance: The roles of maternal control and child temperament. *Journal of Applied* Retrieved from http://www.sciencedirect.com/science/article/pii/S0193397397800081
- Broidy, L. M., Nagin, D. S., Tremblay, R. E., Bates, J. E., Brame, B., Dodge, K. A., ... & Lynam, D. R. (2003). Developmental trajectories of childhood disruptive behaviors and adolescent delinquency: a six-site, cross-national study. *Developmental psychology*, 39(2), 222.
- Brook, M., & Kosson, D. S. (2013). Impaired cognitive empathy in criminal psychopathy: Evidence from a laboratory measure of empathic accuracy.
- Bruner, J. S. (1972). The Uses of Immaturity. *New York University Education Quarterly*, 27(8), 687–708. http://doi.org/10.1037/h0033144
- Burley, P., & GUINNESS, J. M. (1977). Effects of social intelligence on the Milgram paradigm. *Psychological Reports*. Retrieved from http://www.amsciepub.com/doi/pdf/10.2466/pr0.1977.40.3.767
- Carlsmith, J., Lepper, M., & Landauer, T. (1974). Children's obedience to adult requests: Interactive effects of anxiety arousal and apparent punitiveness of the adult. *Journal of Personality and* Retrieved from http://psycnet.apa.org/journals/psp/30/6/822/
- Cohn, J. F., & Campbell, susan b.; Matias, reinaldo; Hopkins, J. (1989). Face-to-Face Interactions of Postpartum Depressed and Nondepressed Mother-Infant Pairs at Two Months. *Developmental Psychology*, 26(1), 15–23. http://doi.org/10.1037/0012-1649.26.1.15
- Crockenberg, S., & Litman, C. (1990). Autonomy as competence in 2-year-olds: Maternal correlates of child defiance, compliance, and self-assertion. *Developmental Psychology*. Retrieved from http://psycnet.apa.org/journals/dev/26/6/961/
- Dadds, M. R., Hawes, D. J., Frost, A. D. J., Vassallo, S., Bunn, P., Hunter, K., & Merz, S. (2009). Learning to "talk the talk": the relationship of psychopathic traits to deficits in empathy across childhood. *Journal of Child Psychology and Psychiatry*, 50(5), 599–606. http://doi.org/10.1111/j.1469-7610.2008.02058.x
- Dadds, M. R., Perry, Y., Hawes, D. J., Merz, S., Riddell, A. C., Haines, D. J., ... Abeygunawardane, A. I. (2006). Attention to the eyes and fear-recognition deficits in child psychopathy. *The British Journal of Psychiatry : The Journal of Mental Science*, 189, 280– 1. http://doi.org/10.1192/bjp.bp.105.018150

- Damon, W. (1977). Measurement and social development. *The Counseling Psychologist*. Retrieved from http://psycnet.apa.org/psycinfo/1978-31504-001
- De Gelder, B. de, Snyder, J., Greve, D., & Hadjikhani, N. K. (2004). Fear fosters flight: A mechanism for fear contagion when perceiving emotion expressed by a whole body. *Proceedings of the National Academy of Sciences of the United States of America (PNAS)*, 101(47), 16701–16706. Retrieved from http://www.narcis.nl/publication/RecordID/oai:wo.uvt.nl:141110
- De Wied, M., Van Boxtel, A., Matthys, W., & Meeus, W. (2012). Verbal, facial and autonomic responses to empathy-eliciting film clips by disruptive male adolescents with high versus low callous-unemotional traits. *Journal of Abnormal Child Psychology*, 40, 211–223. http://doi.org/10.1007/s10802-011-9557-8
- Dienstbier, R. a. (1984). The role of emotion in moral socialization. *Psychology*, *Paper 114*, 484–514. http://doi.org/10.1016/j.tics.2008.09.006
- Dunn, L. (1997). Examiner's manual for the PPVT-III peabody picture vocabulary test: Form IIIA and Form IIIB. Retrieved from https://scholar.google.com/scholar?q=Dunn+%26+Dunn%2C+1997+ppvt&btnG=&hl=en& as_sdt=0%2C23#0
- Durbin, C. (2010). Validity of young children's self-reports of their emotion in response to structured laboratory tasks. *Emotion*. Retrieved from http://psycnet.apa.org/journals/emo/10/4/519/
- Durbin, C., Hayden, E., Klein, D., & Olino, T. (2007). Stability of laboratory-assessed temperamental emotionality traits from ages 3 to 7. *Emotion*. Retrieved from http://psycnet.apa.org/journals/emo/7/2/388/
- Dyson, M.W., Olino, T.M., Durbin, C.E., Goldsmith, H.H., & Klein, D.N. (2012). The structure of temperament in preschoolers: A two-stage factor analytic approach. *Emotion*, 12(1), 44-57.
- Eisenberg, N., Eggum, N. D., & Edwards, A. (2010). Empathy-related responding and moral development.
- Eisenberg, N., & Lennon, R. (1983). Sex differences in empathy and related capacities. *Psychological Bulletin*. Retrieved from http://psycnet.apa.org/journals/bul/94/1/100/
- Eisenberg, N., & Lundy, T. (1985). Children's justifications for their adult and peer-directed compliant (prosocial and nonprosocial) behaviors. *Developmental Psychology*. Retrieved from http://psycnet.apa.org/journals/dev/21/2/325/

- Eisenberg, N., & Shell, R. (1987). Prosocial development in middle childhood: A longitudinal study. *Developmental* ..., 2(5), 712–718. Retrieved from http://psycnet.apa.org/journals/dev/23/5/712/
- Elms, A., & Milgram, S. (1966). PERSONALITY CHARACTERISTICS ASSOCIATED WITH OBEDIENCE AND DEFIANCE TOWARD AUTHORITATIVE COMMAND. *Journal of Experimental Research in* Retrieved from http://psycnet.apa.org/psycinfo/1967-04552-001
- Epstein, S., & O'Brien, E. (1985). The person–situation debate in historical and current perspective. *Psychological Bulletin*. Retrieved from http://psycnet.apa.org/journals/bul/98/3/513/
- Fairchild, G., Van Goozen, S. H. M., Calder, A. J., Stollery, S. J., & Goodyer, I. M. (2009). Deficits in facial expression recognition in male adolescents with early-onset or adolescence-onset conduct disorder. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, 50(5), 627–36. http://doi.org/10.1111/j.1469-7610.2008.02020.x
- Field, T. M., Woodson, R., Greenberg, R., & Cohen, D. (1982). Discrimination and imitation of facial expression by neonates. *Science (New York, N.Y.)*, 218(4568), 179–81. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/7123230
- Forman, *D.*, & Kochanska, G. (2001). Viewing imitation as child responsiveness: a link between teaching and discipline domains of socialization. *Developmental Psychology*. Retrieved from http://psycnet.apa.org/journals/dev/37/2/198/
- Fowler, P. J., Tompsett, C. J., Braciszewski, J. M., Jacques-Tiura, A. J., & Baltes, B. B. (2009). Community violence: a meta-analysis on the effect of exposure and mental health outcomes of children and adolescents. *Development and Psychopathology*, 21(1), 227–59. http://doi.org/10.1017/S0954579409000145
- Frick, P. J., & Ellis, M. (1999). Callous-Unemotional Traits and Subtypes of Conduct Disorder. *Clinical Child and Family Psychology Review*, 2(3), 149–168. http://doi.org/10.1023/A:1021803005547
- Frick, P. J., & Morris, A. S. (2004). Temperament and developmental pathways to conduct problems. *Journal of Clinical Child and Adolescent Psychology : The Official Journal for the Society of Clinical Child and Adolescent Psychology, American Psychological Association, Division 53, 33*(1), 54–68. http://doi.org/10.1207/S15374424JCCP3301_6
- Frick, P. J., Ray, J. V, Thornton, L. C., & Kahn, R. E. (2014). Can callous-unemotional traits enhance the understanding, diagnosis, and treatment of serious conduct problems in children and adolescents? A comprehensive review. *Psychological Bulletin*, 140(1), 1–57. http://doi.org/10.1037/a0033076

- Frick, P., & Kimonis, ER, Dandreaux, DM, Farrell, J. (2003). The 4 year stability of psychopathic traits in non-referred youth. *Behavioral Sciences & ..., 21*(6), 713–736. Retrieved from http://onlinelibrary.wiley.com/doi/10.1002/bsl.568/full
- Frick, P., & White, S. (2008). Research review: The importance of callous-unemotional traits for developmental models of aggressive and antisocial behavior. *Journal of Child Psychology and Psychiatry*. Retrieved from http://onlinelibrary.wiley.com/doi/10.1111/j.1469-7610.2007.01862.x/full
- Frye, D., Zelazo, P., & Palfai, T. (1995). Theory of mind and rule-based reasoning. Cognitive Development. Retrieved from http://www.sciencedirect.com/science/article/pii/0885201495900241
- Gaertner, S. L. (1976). Situational determinants of hurting and helping behavior. Social psychology: An introduction, 111-142.
- Garaigordobil, M. (2009). A comparative analysis of empathy in childhood and adolescence: Gender differences and associated socio-emotional variables. *International Journal of Psychology and Psychological* Retrieved from http://www.sc.ehu.es/ptwgalam/art_completo/2009/IJPPT Empatia 2009.pdf
- Gilligan, C., & Attanucci, J. (1988). Two moral orientations: Gender differences and similarities. *Merrill-Palmer Quarterly* (1982-), 223-237.
- Goldsmith, H., Reilly, J., & Lemery, K. (1995). Preschool Laboratory Temperament Assessment Battery. ... Instrument, University of Retrieved from https://scholar.google.com/scholar?q=Goldsmith%2C+Reilly%2C+Lemery%2C+Longley% 2C+%26+Prescott%2C+1995&btnG=&hl=en&as_sdt=0%2C23#0
- Goldsmith, H., Reilly, J., Lemery, K., & Prescott, A. (2001) The Laboratory Temperament Assessment Battery: Middle Childhood Version. *Unpublished assessment manual.*
- Goldsmith, H., & Rothbart, M. (1993). The laboratory temperament assessment battery (LAB-TAB). University of Wisconsin. Retrieved from https://scholar.google.com/scholar?q=goldsmith+1993+labtab&btnG=&hl=en&as_sdt=0%2C23#0
- Goldsmith, H., & Rothbart, M. (1996). The laboratory temperament assessment battery. *Madison: University of Wisconsin,* Retrieved from http://www.waisman.wisc.edu/twinresearch/researchers/LabTAB- Middle Childhood.pdf
- Hamlin, J. K., Wynn, K., & Bloom, P. (2010). Three-month-olds show a negativity bias in their social evaluations. *Developmental Science*. Retrieved from http://onlinelibrary.wiley.com/doi/10.1111/j.1467-7687.2010.00951.x/full

Hartshorne, H., & May, M. (1928). 1930. *Studies in the Nature of Character*. Retrieved from https://scholar.google.com/scholar?q=Hartshorne+%26+May%2C+1928&btnG=&hl=en&a s_sdt=0%2C23#0

Hastings, P. D., Zahn-Waxler, C., Robinson, J., Usher, B., & Bridges, D. (2000). The development of concern for others in children with behavior problems. *Developmental Psychology*, 36(5), 531–46. Retrieved from http://www.safetylit.org/citations/index.php?fuseaction=citations.viewdetails&citationIds[] =citjournalarticle_270616_38

- Hayden, E. P., Klein, D. N., Durbin, C. E., & Olino, T. M. (2006). Positive emotionality at age 3 predicts cognitive styles in 7-year-old children.Development and Psychopathology, 18(02), 409-423.
- Higbee, K. (1979). Factors affecting obedience in preschool children. *The Journal of Genetic Psychology*. Retrieved from http://www.tandfonline.com/doi/pdf/10.1080/00221325.1979.10534059
- Hoffman, M. (1977). Sex differences in empathy and related behaviors. *Psychological Bulletin*. Retrieved from http://psycnet.apa.org/journals/bul/84/4/712/
- Hoffman, M. L. (1975). Moral internalization, parental power, and the nature of parent-child interaction. *Developmental Psychology*, *11*(2), 228–239. http://doi.org/10.1037/h0076463
- Hoffman, M. L. (1982). Development of prosocial motivation: Empathy and guilt. In *The development of prosocial behavior* (pp. 218–231).
- Holland, R. (1967). Representation of dielectric, elastic, and piezoelectric losses by complex coefficients. *Sonics and Ultrasonics, IEEE Transactions on*. Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=1538402
- Izard, C., Dougherty, L., & Hembree, E. (1983). A system for identifying affect expressions by holistic judgments (AFFEX). Retrieved from https://scholar.google.com/scholar?q=Izard%2C+Dougherty%2C+%26+Hebree%2C+1983 &btnG=&hl=en&as_sdt=0%2C23#0
- Jolliffe, D., & Farrington, D. P. (2004). Empathy and offending: A systematic review and metaanalysis. Aggression and Violent Behavior. http://doi.org/10.1016/j.avb.2003.03.001
- Jr, L. B., & Simpson, J. (2009). The power of the situation: The impact of Milgram's obedience studies on personality and social psychology. *American Psychologist*. Retrieved from http://psycnet.apa.org/journals/amp/64/1/12/
- Kagan, J., Reznick, J., & Snidman, N. (1987). The physiology and psychology of behavioral inhibition in children. *Child Development*. Retrieved from http://www.jstor.org/stable/1130685

- Kagan, J., Snidman, N., Arcus, D., & Reznick, J. (1994). Galen's prophecy: Temperament in human nature. Retrieved from http://psycnet.apa.org/psycinfo/1994-97715-000
- Keenan, K., & Shaw, D. S. (1994). The development of aggression in toddlers: A study of lowincome families. *Journal of abnormal child psychology*, 22(1), 53-77.
- Kimonis, E. R., Frick, *P. J.*, Fazekas, H., & Loney, B. R. (2006). Psychopathy, aggression, and the processing of emotional stimuli in non-referred girls and boys. *Behavioral Sciences & the Law*, 24(1), 21–37. http://doi.org/10.1002/bsl.668
- Kimonis, E. R., Frick, P. J., Skeem, J. L., Marsee, M. A., Cruise, K., Munoz, L. C., ... & Morris, A. S. (2008). Assessing callous–unemotional traits in adolescent offenders: Validation of the Inventory of Callous–Unemotional Traits. *International journal of law and psychiatry*, 31(3), 241-252.
- Kochanska, G. (1991). Socialization and Temperament in the Development of Guilt and Conscience. *Child Development*, 62(6), 1379–1392. http://doi.org/10.1111/j.1467-8624.1991.tb01612.x
- Kochanska, G. (1993). Toward a synthesis of parental socialization and child temperament in early development of conscience. *Child Development*. Retrieved from http://onlinelibrary.wiley.com/doi/10.1111/j.1467-8624.1993.tb02913.x/abstract
- Kochanska, G. (1993). Toward a Synthesis of Parental Socialization and Child Temperament in Early Development of Conscience. *Child Development*, *64*(2), 325–347. http://doi.org/10.1111/j.1467-8624.1993.tb02913.x
- Kochanska, G. (2002). Mutually Responsive Orientation Between Mothers and Their Young Children: A Context for the Early Development of Conscience. *Current Directions in Psychological Science*, *11*(6), 191–195. http://doi.org/10.1111/1467-8721.00198
- Kochanska, G., & Aksan, N. (2004). Conscience in Childhood: Past, Present, and Future. *Merrill-Palmer Quarterly*, 50(3), 299–310. http://doi.org/10.1353/mpq.2004.0020
- Kochanska, G., & Aksan, N. (2006). Children's conscience and self-regulation. *Journal of Personality*, 74(6), 1587–617. http://doi.org/10.1111/j.1467-6494.2006.00421.x
- Kochanska, G., Aksan, N., & Koenig, A. (1995). A longitudinal study of the roots of preschoolers' conscience: Committed compliance and emerging internalization. *Child Development*. Retrieved from http://onlinelibrary.wiley.com/doi/10.1111/j.1467-8624.1995.tb00963.x/abstract
- Kochanska, G., Aksan, N., & Nichols, K. (2003). Maternal power assertion in discipline and moral discourse contexts: commonalities, differences, and implications for children's moral conduct and cognition. *Developmental Psychology*. Retrieved from http://psycnet.apa.org/journals/dev/39/6/949/

- Kochanska, G., Coy, K. C., & Murray, K. T. (2001). The Development of Self-Regulation in the First Four Years of Life. *Child Development*, 72(4), 1091–1111. http://doi.org/10.1111/1467-8624.00336
- Kochanska, G., DeVet, K., Goldman, M., Murray, K., & Putnam, S. P. (1994). Maternal Reports of Conscience Development and Temperament in Young Children. *Child Development*, 65(3), 852–868. http://doi.org/10.1111/j.1467-8624.1994.tb00788.x
- Kochanska, G., & Forman, D. (2005). Pathways to conscience: Early mother–child mutually responsive orientation and children's moral emotion, conduct, and cognition. *Journal of Child* Retrieved from http://onlinelibrary.wiley.com/doi/10.1111/j.1469-7610.2004.00348.x/full
- Kochanska, G., Gross, J. N., Lin, M.-H., & Nichols, K. E. (2002). Guilt in Young Children: Development, Determinants, and Relations with a Broader System of Standards. *Child Development*, 73(2), 461–482. http://doi.org/10.1111/1467-8624.00418
- Kochanska, G., & Knaack, A. (2003). Effortful control as a personality characteristic of young children: Antecedents, correlates, and consequences. *Journal of Personality*. Retrieved from http://onlinelibrary.wiley.com/doi/10.1111/1467-6494.7106008/full
- Kochanska, G., Murray, K., & Coy, K. (1997). Inhibitory control as a contributor to conscience in childhood: From toddler to early school age. Child Development. Retrieved from http://onlinelibrary.wiley.com/doi/10.1111/j.1467-8624.1997.tb01939.x/full
- Kochanska, G., Murray, K., Jacques, T. Y., Koenig, A. L., & Vandegeest, K. A. (1996). Inhibitory control in young children and its role in emerging internalization. Child development, 67(2), 490-507.
- Kochanska, G., Padavich, D. L., & Koenig, A. L. (1996). Children's Narratives about Hypothetical Moral Dilemmas and Objective Measures of Their Conscience: Mutual Relations and Socialization Antecedents. *Child Development*, 67(4), 1420–1436. http://doi.org/10.1111/j.1467-8624.1996.tb01805.x
- Kochanska, G., Tjebkes, J., & Fortnan, D. (1998). Children's emerging regulation of conduct: Restraint, compliance, and internalization from infancy to the second year. Child Development. Retrieved from http://onlinelibrary.wiley.com/doi/10.1111/j.1467-8624.1998.tb06218.x/abstract
- Kohlberg, L., & Hersh, R. (1977). Moral development: A review of the theory. *Theory into Practice*. Retrieved from http://www.tandfonline.com/doi/pdf/10.1080/00405847709542675
- Kohlberg, L., & Kramer, R. (1969). Continuities and discontinuities in childhood and adult moral development. *Human Development*. Retrieved from http://www.karger.com/Article/Abstract/270857

- Kolko, D. J., & Kazdin, A. E. (1993). Emotional/behavioral problems in clinic and nonclinic children: Correspondence among child, parent and teacher reports. *Journal of Child Psychology and Psychiatry*, 34(6), 991-1006.
- Krahn, J. (1971). A comparison of Kohlberg's and Piaget's type I morality. *Religious Education*. Retrieved from http://www.tandfonline.com/doi/pdf/10.1080/0034408710660509
- Krettenauer, T., Asendorpf, J. B., & Nunner-Winkler, G. (2013). Moral emotion attributions and personality traits as long-term predictors of antisocial conduct in early adulthood: Findings from a 20-year longitudinal study. *International Journal of Behavioral Development*, 37, 192–201. http://doi.org/10.1177/0165025412472409
- Kuczynski, L. (1987). A developmental interpretation of young children's noncompliance. *Developmental* Retrieved from http://psycnet.apa.org/journals/dev/23/6/799/
- Kuczynski, L., & Kochanska, G. (1990). Development of children's noncompliance strategies from toddlerhood to age 5. *Developmental Psychology*. Retrieved from http://psycnet.apa.org/journals/dev/26/3/398/
- Kuczynski, L., Kochanska, G., Radke-Yarrow, M., & Girnius-Brown, O. (1987). A developmental interpretation of young children's noncompliance.Developmental Psychology, 23(6), 799.
- Landauer, T., Carlsmith, J., & Lepper, M. (1970). Experimental analysis of the factors determining obedience of four-year-old children to adult females. *Child Development*. Retrieved from http://www.jstor.org/stable/1127210
- Laupa, M. (1991). Children's reasoning about three authority attributes: Adult status, knowledge, and social position. *Developmental Psychology*. Retrieved from http://psycnet.apa.org/journals/dev/27/2/321/
- Laupa, M. (1994). "Who's in charge?" Preschool children's concepts of authority. *Early Childhood Research Quarterly*. Retrieved from http://www.sciencedirect.com/science/article/pii/0885200694900264
- Laupa, M., & Turiel, E. (1986). Children's conceptions of adult and peer authority. *Child Development*. Retrieved from http://www.jstor.org/stable/1130596
- Lickenbrock, D. (2013). Early temperament and attachment security with mothers and fathers as predictors of toddler compliance and noncompliance. *Infant and Child* Retrieved from http://onlinelibrary.wiley.com/doi/10.1002/icd.1808/full
- Loney, B. R., Frick, P. J., Clements, C. B., Ellis, M. L., & Kerlin, K. (2003). Callousunemotional traits, impulsivity, and emotional processing in adolescents with antisocial behavior problems. *Journal of Clinical Child and Adolescent Psychology : The Official Journal for the Society of Clinical Child and Adolescent Psychology, American*

Psychological Association, Division 53, 32(1), 66–80. http://doi.org/10.1207/S15374424JCCP3201_07

- Lovett, B. J., & Sheffield, R. A. (2007). Affective empathy deficits in aggressive children and adolescents: a critical review. *Clinical Psychology Review*, 27(1), 1–13. http://doi.org/10.1016/j.cpr.2006.03.003
- Eisenberg, N., Lundy, T., Shell, R., & Roth, K. (1985). Children's justifications for their adult and peer-directed compliant (prosocial and nonprosocial) behaviors. Developmental Psychology, 21(2), 325.
- Marsh, A. A., & Cardinale, E. M. (2014). When psychopathy impairs moral judgments: neural responses during judgments about causing fear. *Social Cognitive and Affective Neuroscience*, 9(1), 3–11. http://doi.org/10.1093/scan/nss097
- Milgram, S. (1961). Nationality and conformity. *Scientific American*. Retrieved from http://psycnet.apa.org/psycinfo/1962-06297-001
- Milgram, S. (1963). Behavioral study of obedience. *The Journal of Abnormal and Social Psychology*. Retrieved from http://psycnet.apa.org/journals/abn/67/4/371/
- Milgram, S. (1965). Some conditions of obedience and disobedience to authority. *Human Relations*. Retrieved from http://psyc604.stasson.org/Milgram2.pdf
- Milgram, S. (1974). The experience of living in cities. *Crowding and Behavior*. Retrieved from https://books.google.com/books?hl=en&lr=&id=XA8Ho_aITVoC&oi=fnd&pg=PA41&dq= milgram+1974&ots=JwPHwRr9jF&sig=rzyyuqi7vGmYPvZDAXtrMqxRuD4
- Mixon, D. (1976). Studying feignable behavior. *Representative Research in Social Psychology*. Retrieved from http://psycnet.apa.org/psycinfo/1977-29451-001
- Muñoz, L. C. (2009). Callous-unemotional traits are related to combined deficits in recognizing afraid faces and body poses. *Journal of the American Academy of Child and Adolescent Psychiatry*, 48(5), 554–62. http://doi.org/10.1097/CHI.0b013e31819c2419
- Piaget, J. (1932). The moral development of the child. *Kegan Paul, London*. Retrieved from https://scholar.google.com/scholar?q=piaget+1932&btnG=&hl=en&as_sdt=0%2C23#1
- Piaget, J. (1965). The moral judgment of the child (1932). *New York: The Free*. Retrieved from http://scholar.google.com/scholar?q=piaget+1965+moral+judgement+of+the+child&btnG= &hl=en&as_sdt=0,23#4
- Pons, F., Harris, P. L., & de Rosnay, M. (2004). Emotion comprehension between 3 and 11 years: Developmental periods and hierarchical organization. *European Journal of Developmental Psychology*, 1(2), 127–152. http://doi.org/10.1080/17405620344000022

- Rose, S. R. (1999). Towards the Development of an Internalized Conscience. Journal of Human Behavior in the Social Environment, 2(3), 15–28. http://doi.org/10.1300/J137v02n03_02
- Rose, S. L., Rose, S. A., & Feldman, J. F. (1989). Stability of behavior problems in very young children. *Development and Psychopathology*, *1*(01), 5-19.
- Rose, A. J., & Rudolph, K. D. (2006). A review of sex differences in peer relationship processes: potential trade-offs for the emotional and behavioral development of girls and boys. *Psychological bulletin*, *132*(1), 98.
- Ross, L. (1977). The intuitive psychologist and his shortcomings: Distortions in the attribution process. Advances in experimental social psychology, 10, 173-220.
- Rothbart, M., Ahadi, S., & Hershey, K. (1994). Temperament and social behavior in childhood. *Merrill-Palmer Quarterly* (1982-). Retrieved from http://www.jstor.org/stable/23087906
- Rothbart, M., & Bates, J. (1998). Temperament. Hoboken. Retrieved from https://scholar.google.com/scholar?q=Rothbart+%26+Bates%2C+1998&hl=en&as_sdt=0% 2C23&as_ylo=1998&as_yhi=1998#0
- Rothbart, M. K., Ahadi, S. A., Hershey, K. L., & Fisher, P. (2001). Investigations of Temperament at Three to Seven Years: The Children's Behavior Questionnaire. *Child Development*, 72(5), 1394–1408. http://doi.org/10.1111/1467-8624.00355
- Roth-Hanania, R., Davidov, M., & Zahn-Waxler, C. (2011). Empathy development from 8 to 16 months: early signs of concern for others. *Infant Behavior & Development*, 34(3), 447–58. http://doi.org/10.1016/j.infbeh.2011.04.007
- Sagi, A., & Hoffman, M. (1976). Empathic distress in the newborn. *Developmental Psychology*, *12*(2), 175–176. http://doi.org/10.1037/0012-1649.12.2.175
- Shalala, S. R. (1975). A study of various communication settings which produce obedience by subordinates to unlawful superior orders (Doctoral dissertation, ProQuest Information & Learning).
- Shanab, M., & Yahya, K. (1977). A behavioral study of obedience in children. Journal of Personality and Social Retrieved from http://psycnet.apa.org/journals/psp/35/7/530/
- Shelton, K. K., Frick, P. J., & Wootton, J. (1996). Assessment of parenting practices in families of elementary school-age children. *Journal of clinical child psychology*, 25(3), 317-329.
- Shimizu, Y. A., & Johnson, S. C. (2004). Infants' attribution of a goal to a morphologically unfamiliar agent. *Developmental Science*, 7(4), 425–430. http://doi.org/10.1111/j.1467-7687.2004.00362.x

- Snyder, M., & Ickes, W. (1985). Personality and social behavior. *Handbook of Social Psychology*. Retrieved from https://scholar.google.com/scholar?q=Snyder+%26+Ickes%2C+1985&btnG=&hl=en&as_s dt=0%2C23#0
- Stanger, C., & Lewis, M. (1993). Agreement among parents, teachers, and children on internalizing and externalizing behavior problems. *Journal of Clinical Child Psychology*, 22(1), 107-116.
- Stifter, C. A., Cipriano, E., Conway, A., & Kelleher, R. (2009). Temperament and the Development of Conscience: The Moderating Role of Effortful Control. *Social Development (Oxford, England)*, 18(2), 353–374. http://doi.org/10.1111/j.1467-9507.2008.00491.x
- Strommen, E. (1973). Verbal Self-Regulation in a Children's Game: Impulsive Errors on" Simon Says." *Child Development*. Retrieved from http://www.jstor.org/stable/1127737
- Tan, P., Steinbach, M., & Kumar, V. (2006) *Cluster Analysis: Basic Concepts and Algorithms*. Retrieved from http://www-users.cs.umn.edu/~kumar/dmbook/ch8.pdf
- Tibshirani, R., Walther, G., & Hastie, T. (2001). Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(2), 411-423.
- Vroman, L., Lo, S., & Durbin, C. (2014). Structure and convergent validity of children's temperament traits as assessed by experimenter ratings of child behavior. *Journal of Research in Personality*. Retrieved from http://www.sciencedirect.com/science/article/pii/S0092656614000531
- Willoughby, M. T., Mills-Koonce, W. R., Gottfredson, N. C., & Wagner, N. (2014). Measuring Callous Unemotional Behaviors in Early Childhood: Factor Structure and the Prediction of Stable Aggression in Middle Childhood. *Journal of Psychopathology and Behavioral Assessment*, 36(1), 30–42. http://doi.org/10.1007/s10862-013-9379-9
- Willoughby, M. T., Waschbusch, D. A., Moore, G. A., & Propper, C. B. (2011). Using the ASEBA to Screen for Callous Unemotional Traits in Early Childhood: Factor Structure, Temporal Stability, and Utility. *Journal of Psychopathology and Behavioral Assessment*, 33(1), 19–30. http://doi.org/10.1007/s10862-010-9195-4
- Wilson, S., & Durbin, C.E. (2012). Dyadic parent-child interaction during early childhood: Contributions of parental and child personality traits. *Journal of Personality*, 80(5), 1313-1338.
- Zahn-Waxler, C., Radke-Yarrow, M., Wagner, E., & Chapman, M. (1992). Development of concern for others. *Developmental Psychology*, *28*(1), 126–136. Retrieved from http://cat.inist.fr/?aModele=afficheN&cpsidt=5053281

Zelazo, *P.*, Müller, U., & Frye, *D.* (2003). The development of executive function in early childhood. *Monographs of the* Retrieved from http://www.jstor.org/stable/1166202

Zimbardo, P. G. (1974). On" Obedience to authority.".