Journal of Experimental Criminology (2023) 19:713–742
https://doi.org/10.1007/s11292-021-09496-5

The effects of spanking on psychosocial outcomes: revisiting genetic and environmental covariation

Nicole Barbaro[1] · Eric J. Connolly[2] · Madi Sogge[3] · Todd K. Shackelford[1] · Brian B. Boutwell[4,5]

Accepted: 22 November 2021 / Published online: 18 April 2022

Abstract

Background There is a vast literature on the negative associations between spanking in childhood and various psychosocial developmental outcomes; yet, control for potential genetic confounds is rare.

Objectives The current research aimed to provide probable ranges of estimates of the degree to which genetic and nonshared environmental covariation could explain the reported phenotypic effects in the Gershoff and Grogan-Kaylor (Gershoff and Grogan-Kaylor, Family Relations 65:490–501, 2016a, Gershoff and Grogan-Kaylor, Journal of Family Psychology 30:453, 2016b) meta-analysis of spanking.

Participants and setting.

The analytic sample for Study 1 was secured from the Children of the National Longitudinal Survey of Youth (CNLSY) and consisted of 2868 respondents (siblings and half-siblings). The data for Study 2 were secured from the published literature.

Methods Study 1 analyzed the data from the CNLSY using univariate ACE models and bivariate Cholesky decomposition models. Study 2 used simulation modeling to provide a summative evaluation of the psychosocial effects of spanking with regard to genetic and nonshared environmental covariation.

Results Study 1 replicated previous work showing that associations between spanking and outcomes of delinquency, depression, and alcohol use were explained by moderate-to-large degrees of genetic covariation and small-to-moderate degrees of nonshared environmental covariation. Simulation estimates from Study 2 suggest that genetic covariation accounts for a substantial amount of the phenotypic effect between spanking and psychosocial outcomes (≈60–80%), with the remainder attributable to nonshared environmental covariation (≈0–40%).

Conclusions Results of the current research indicate that continued work on the effects of spanking is best served by behavior genetic research on a broader range of outcomes than what is currently available.

Keywords Spanking · Genetic covariation · Environmental covariation · Simulation modeling · Behavioral genetics

Extended author information available on the last page of the article

The extent to which the experiences of children with their parents or primary caregivers resonant across the decades of the life course has long been a central focus for developmental researchers (Durrant & Ensom, 2012). The parental use of spanking, in particular, defined as "hitting a child on their buttocks or extremities using an open-hand" (Gershoff & Grogan-Kaylor, 2016a, b, p. 453) has received substantial empirical attention within the developmental sciences (Gershoff, 2002; Gershoff & Grogan-Kaylor, 2016a, b; Larzelere & Kuhn, 2005; Paolucci & Violato, 2004).[1] Paradoxically, though a goal among those using spanking might be to promote prosocial growth in children, the use of spanking has been consistently linked with internalizing and externalizing problems in childhood and adolescence (e.g., Lansford et al., 2012), mental health problems in adulthood (e.g., Afifi et al., 2017), and a range of other adverse outcomes across the life course (see Gershoff & Grogan-Kaylor, 2016a, b).

Although substantial empirical effort has been devoted to elucidating the enduring outcomes of spanking children, meta-analyses of this body of work have reached somewhat varying conclusions. Results differ, for instance, regarding whether spanking actually reduces any form of problem behaviors (e.g., Gershoff, 2002; Larzelere & Kuhn, 2005). The precise degree to which spanking is linked with various negative psychosocial outcomes (e.g., Ferguson, 2013; Paolucci & Violato, 2004), moreover, also differs across studies. Gershoff and Grogan-Kaylor (2016a, b) provided what is currently the most recent meta-analytic addition to the spanking literature in an effort to address some of the lingering questions on this topic. Methodologically important, both here in and in the 2016 meta-analysis, was ensuring that spanking (as defined above) was not confounded with harsher forms of physical punishment perpetrated against children (e.g., hitting children with objects; abusive behaviors including choking and beating — "physical abuse").

This distinction is important and warrants additional consideration given the broader implications not only for the current study, but also for all fields examining issues of child development. The most pressing point to make, perhaps, is that a consensus already exists among scholars regarding the impact of *physical abuse*, which is detrimental to healthy child development (see Ferguson, 2013; Jaffee et al., 2004; Larzelere & Kuhn, 2005; Lynch et al., 2006). The risks of abuse and neglect include adverse outcomes such as physical injury, dramatically diminished health, and increased disease risk, as well as concomitant and prolonged emotional, behavioral, and psychiatric problems (Danese & Tan, 2014; Infurna et al., 2016). Yet, the methodological and theoretical distinction between the two—abuse versus more normative punishment—should be attended by researchers interested in these issues for at least two reasons.

First, researchers make distinctions between the two to avoid conflating the effects of one variable with the effects of the other (Lynch et al., 2006). Second, and relatedly, it is plausible that *both* spanking and abuse are causally linked to adverse developmental outcomes. Failure to distinguish between the two thus makes

[1] "Spanking," "corporal punishment," and "mild physical punishment" are viewed largely as synonymous. Here, we maintain the use of the term "spanking" for consistency throughout, as defined above.

determinations along these lines difficult or impossible. For this and other reasons, moreover, it continues to remain unclear whether spanking, relative to abuse, is: (1) inversely *correlated* with negative outcomes (Grogan-Kaylor et al., 2018), or (2) *causes* negative psychosocial outcomes throughout the life course (Cecil et al., 2012; Lynch et al., 2006).

One of the more recent and thorough attempts to systematically assess the state of the literature was by Gershoff and Grogan-Kaylor (2016a, 2016b). Their results revealed that spanking was, indeed, associated with wide-ranging negative psychological and behavioral outcomes across development. Psychosocial outcomes linked to spanking vs. non-spanking, importantly, did not differ as a function of study design (e.g., retrospective, longitudinal), measurement (e.g., observational, child-report, parent-report), or age of children at the time of spanking (ranging from toddlerhood to adolescence). Taken together, the results of Gershoff and Grogan-Kaylor align with the prior consensus that spanking is associated with poorer developmental outcomes, appears to be ineffective at reducing problem behaviors, and has failed to produce any net-positive effects on child development (Durrant & Ensom, 2012; Gershoff, 2013; Grogan-Kaylor et al., 2018).

Despite robust associations between spanking and adverse outcomes (Gershoff & Grogan-Kaylor, 2016a, b), methodological limitations of prior work advise against making stronger forms of *causal* inference (see Lee, 2012). To understand why, one must first recall that for ethical and practical reasons, research on spanking is largely observational (retrospective reporting, cross-sectional, longitudinal; more than 90% of studies included in Gershoff & Grogan-Kaylor, 2016a, b) rather than experimental. Like much social science research, scholars must rely on survey methods and statistical techniques when attempting to parse selection effects from the purported causal effects of spanking (see Gershoff et al., 2018; Lee, 2012). This is not a criticism, but instead simply reflects a reality when studying this and related topics in the behavioral sciences.

In such instances, what is necessary is a concerted effort to correct for biases when examining exposure effects that are not randomized (Lee, 2012). For many studies in the behavioral sciences, in general, and those examining spanking in particular, some variant of multiple regression is typically used in an attempt to parse the effects of possible confounders from the primary variable of interest. Such approaches can be useful for causal modeling assuming certain assumptions are reliably met (Lee, 2012). What makes this difficult is the fact that the most ardent attempts to subvert confounding—via inclusion of numerous control variables—can, paradoxically, produce results that are less, not more, precise (see Lee, 2012; Rohrer, 2018). Reflecting broadly on this topic, Lee (2012) has urged a deeper appreciation of at least two points. First, approaches using statistical control as opposed to experimental presume a deep knowledge of, and access to, the variables that must be *included* in an equation. Second, such approaches also presume similarly deep insight about those variables which ought to be *excluded* (Lee, 2012). Satisfying one of these conditions, much less both, can frequently constitute an undue expectational burden for any researcher.

Adjacent to this methodological concern is a broader philosophical point about a common approach to causation and causal inference in the social sciences (Holland,

1986), especially in relation to research topics like spanking (for expanded definitions of "cause" that may be preferred, see Lee, 2012). The association between spanking and adverse psychosocial outcomes is typically inferred to be *consistent with* a causal interpretation given that (1) correlational research documents significant, reliable associations between parental use of spanking and adverse psychosocial outcomes across development, (2) spanking precedes (rather than follows) adverse psychosocial outcomes, and (3) alternative, potentially confounding variables are ruled out via statistical controls (see Grogan-Kaylor et al., 2018). Because empirical work has generated relatively robust support for these criteria (see Gershoff et al., 2018), the consensus is that spanking is *consistent with* causal interpretations, such that spanking has a direct, negative impact on psychosocial outcomes—acknowledging, of course, that indisputable causal conclusions are not tenable (see Gershoff & Grogan-Kaylor, 2016a, b; Lynch et al., 2006).

Our concerns here regarding the claim of a direct association between spanking and adverse outcomes centers on the third criterion. In particular, it is not presently clear that relevant confounds have been definitively ruled out and, moreover, some control variables which should have been excluded from modeling strategies may have been included (Lee, 2012). Criticisms of this nature fall under the umbrella of selection effects and omitted variables bias (for other possibilities, see Lee, 2012). To begin illustrating this point more concretely, consider first the issue that while a large proportion of Americans report using spanking (Gershoff & Grogan-Kaylor, 2016a, b), there is still considerable variation in who spanks, how frequently they spank, when spanking is used, and how aggressively spanking is inflicted (Benjet & Kazdin, 2003).

Informing this variation are factors such as caregiver personality traits, demographic variables, and a range of other factors influencing parenting strategies (including those associated with spanking; e.g., Gershoff & Grogan-Kaylor, 2016a, b). Children also vary in regard to factors that influence parenting strategies and spanking (Cecil et al., 2012; Larzelere et al., 2004)—known as evocative child-driven effects (Scarr & McCartney, 1983). Finally, parents typically provide other broader aspects of rearing environments for their children (in addition to their genes), which may further impact the child's behavior—known as passive child influences (Scarr & McCartney, 1983). These factors are all capable of introducing confounds which may act to attenuate causal effects of spanking when fully corrected.

When assessing any parenting variable on a child outcome, there is also the issue of gene-environment correlation (eluded to above), which may account for part of an otherwise robust association between parental behavior and child outcome (Scarr & McCartney, 1993). A gene-environment correlation refers to the association between partly heritable phenotypic characteristics and nonrandom exposure to environmental experiences that are associated with that same phenotypic characteristic (Plomin et al., 1977). Put differently, genotypes are often non-randomly associated with environment, via manifested behaviors and psychological traits of caregivers and children (Kandler & Zapko-Willmes, 2017). Put differently, the environment might be considered to be an extension of an individual's cognitive phenotypes, which is influenced further upstream by corresponding genotypes (Plomin,

1977; Scarr & McCartney, 1993). Owing to these processes, moreover, the problem of genetic covariation between parenting and child outcomes can lurk in parenting studies. Genetic covariation refers to the problem in which two variables that covary at the phenotypic level—such as spanking and psychosocial outcomes—are partly heritable and, therefore, may also share genetic architectures (Boutwell & Adams, 2020). The implication is that associations between some variables at the phenotypic level may be partly a result of a third variable—shared genetic factors—underpinning the phenotypic traits.

The term "heritable" refers to traits for which a proportion of phenotypic variation is explained by genetic variation (Polderman et al., 2015). Nearly 50 years of research has provided a robust cannon of findings concerning the relative contribution of genes and environment to development, which one might summarize thusly (see Chabris et al., 2015; Plomin et al., 2016; Polderman et al., 2015; Turkheimer, 2000 for lengthier original discussions): (1) virtually all complex traits are to varying degrees heritable; (2) similarities between biological relatives are often largely the result of shared genetic variation; (3) environmental experiences unique to siblings raised together consistently capture variation in developmental outcomes, and (4) genetic effects on complex traits are polygenic and pleiotropic in nature, meaning that each trait is influenced by many genes, and that a given gene can be expected to associate with multiple traits and exert relatively miniscule effects on that trait (Chabris et al., 2014). An implication of these "laws" is that, to the extent that two heritable traits covary at the genetic level, genetic covariation partly explaining phenotypic associations is a plausible possibility which should be dealt with empirically (Barnes et al., 2014a). Just as researchers include statistical control variables in regression-based analyses to identify unique associations between variables, if known genetic variation underpinning traits of interest is unaccounted for, then it is possible that a significant phenotypic association appears despite the fact that the phenotypic association may largely be explained by genetic covariation (Lee, 2012).

Complicating the matter is the fact that data produced via standard social science research designs, such as longitudinal designs or retrospective reports, often include assessments from only one child per family and, therefore, cannot account for potential genetic covariation between traits of interest (Barnes et al., 2014a, 2014b). The only direct way to account for genetic covariation in developmental research (outside of experimentation) is to use a genetically informative research design, which typically means analyzing data from twins or siblings (Plomin et al., 2013). With regard to spanking and consequent outcomes, the variables accord with the behavioral genetic laws, such that each is partly heritable (see, Barnes et al., 2013; Burt, 2009; Polderman et al., 2015), and the traits covary genetically (Barnes et al., 2013; Button et al., 2008; Jaffee et al., 2004); therefore, genetic covariation is not only plausible (Barnes et al., 2013), but also could explain a large proportion of the observed phenotypic association between spanking and psychosocial outcomes (see, Button et al., 2008; Jaffee et al., 2004).

Genetically informative research designs capable of identifying shared genetic covariation between spanking and relevant outcomes have, indeed, supported this possibility. Jaffee et al. (2004) used a sample of twins to investigate whether the association between spanking (referred to as "corporal punishment" in Jaffee

et al., 2004) and children's antisocial behavior was, to any extent, accounted for by genetic covariation between the traits. Jaffee et al. found that genetic covariation between spanking and antisocial behavior accounted for around 86% of the observed phenotypic association, whereas the remainder was primarily explained by non-shared environmental covariation. Another twin study conducted by Button et al. (2008) found that genetic covariation between parental punitive punishment and children's externalizing behaviors explained 61% (for maternal punitive punishment) and 98% (for paternal punitive punishment) of the phenotypic association. Together, the results of these studies are consistent with the notion that associations between spanking behaviors and behavioral outcomes can be at least partially accounted for by shared genetic influences.

While the twin studies by Jaffee et al. (2004) and Button et al. (2008) suggest that genetic covariation can account for some proportion of the phenotypic effect between spanking and negative outcomes, the distinction between spanking and harsher forms of parental punishment still lingers as a concern (see Gershoff & Grogan-Kaylor, 2016a, b). In addition to spanking, Jaffee et al. examined genetic and environmental covariation of the association between harsher parenting (i.e., physical abuse) and children's antisocial behavior. In contrast to the significant genetic overlap found for spanking and antisocial behavior, Jaffee et al. found minimal genetic influences on children's experiences of physical abuse, along with significant and substantial effects of shared environmental factors on such experiences. Similarly, a study by Lynch et al. (2006) using the children-of-twins design investigated effects of mild physical punishment (i.e., spanking) and harsher physical punishment (i.e., physical abuse) on internalizing and externalizing problems, and substance use. The results reported by Lynch et al. revealed that, after controlling for genetic relatedness, spanking had only slight associations with negative outcomes. In contrast, harsh physical punishment did have significant negative effects on outcomes after genetic relatedness was accounted for. Assessed collectively, the results of Jaffee et al. and Lynch et al. suggest that a *causal* relationship likely exists between harsh physical punishment (i.e., something closely akin to abuse) and antisocial behavior. A finding which accords with the conclusion of Gershoff and Gorgon-Kaylor (2016) that proper operationalization of spanking (vs. physical abuse) is necessary to understand the precise effects of spanking on child development (see also Lynch et al., 2006).

In pursuit of this goal, and using cross-lagged longitudinal panel analyses, Cecil et al. (2012) probed a related research question when they examined whether harsh physical punishment (i.e., smacking and shouting) was associated with self-control difficulties from early childhood to adolescence. The results suggested bi-directional effects between harsh punishment and self-control, such that each influenced the other. One caveat, though, was that for children 7–12 years old, only evocative effects were found, such that self-control difficulties predicted later harsh punishment. Harsh punishment did not predict later self-control difficulties, however. Cecil et al. (2012) did find long-term effects of harsh punishment on early adolescent conduct problems, but for only boys. Once again, these results are consistent with the notion that children's individual behavior can

explain variance in the application of parental punishments such as spanking, and thus needs to be accounted for in research designs (Cecil et al., 2012).

Adoption studies also provide insights on this topic, and offer additional support for the relevance of evocative child effects when trying to understand and clarify the effects of physical punishment (Cecil et al., 2012). Research by O'Connor et al. (1998), for example, indicated that genetically at-risk adoptees (i.e., based on biological mother's antisocial behavior scores prior to the birth of children) were more likely to receive negative parenting from adoptive parents as compared to children at lower genetic risk. Subsequent adoption studies accord with these findings, and also extend them by demonstrating the ability of certain environments (Hao & Matsueda, 2006) to exacerbate evocative effects. In particular, genetically at-risk children may react more negatively to harsh parenting than do children at lower genetic risk, thereby magnifying the negative impacts of spanking (Riggins-Caspers et al., 2003).

Collectively, these studies suggest that the effects of spanking are complex and efforts to conceptualize and model associations with child outcomes require nuance. Additionally, prior work is consistent with the conclusion that at least a portion of the direct links between spanking (but not physical abuse) and negative outcomes could arise from genetic covariation between the traits (Button et al., 2008; Jaffee et al., 2004). Mechanistically, this describes an evocative process whereby children's genetic propensities for problem behavior capture some of the variance in punishment technique used by the parent (O'Connor et al., 1998), which in turn has negative effects, particularly in early and middle childhood (Cecil, et al., 2012). The remaining proportion of the association seems to be the result of nonshared environmental effects (Button et al., 2008; Jaffee et al., 2004), which is further consistent with the idea that spanking may have negative causal effects on child development (Gershoff & Grogan-Kaylor, 2016a, b; Lynch et al., 2006). The broader point is that if the magnitude of the enduring effects of spanking is to be accurately estimated, then methodological approaches which can estimate and parse environmental and genetic effects seem necessary (see Briley et al., 2019; Lee, 2012).

With this in mind, our goal is to provide an initial evaluation of the current state of the spanking literature with regard to genetic and (nonshared) environmental covariation. To this end, our starting point is the most up-to-date meta-analysis conducted by Gershoff and Grogan-Kaylor (2016a, b) assessing the phenotypic associations between spanking and a broad range of psychosocial outcomes across the life course. The available behavior genetic research on spanking includes a more limited range of outcomes, such as internalizing and externalizing problems, and antisocial or conduct behavioral problems. In contrast, the broader developmental spanking literature has assessed a wide range of psychosocial outcomes linked to spanking: 17 specific psychosocial outcomes and the "overall effect" of spanking in the recent meta-analysis, to be specific (Gershoff & Grogan-Kaylor, 2016a, b).

To provide an evaluative summary of the literature and groundwork for future behavior genetic research, we calculated a range of estimates reflecting the degree to which genetic and nonshared environmental covariation can explain the reported phenotypic associations reported in Gershoff and Grogan-Kaylor. In addition to using parameters available from the published literature, we also included independently generated estimates from new analyses of data from the

Children of the National Longitudinal Survey of Youth (CNLSY). By obtaining heritability estimates and environmental estimates on spanking and the significant psychosocial outcomes in Gershoff and Grogan-Kaylor, combined with genetic and nonshared environmental correlations provided by Button et al. (2008), Jaffee et al. (2004), and original analyses reported here from the CNLSY, we are able to provide estimates of genetic and nonshared environmental covariation for all significant phenotypic effects.

Study 1

Study 1 reports the results of new analyses on data from the CNLSY. The study assesses relative genetic and environmental contributions to the association between maternal reports of spanking and the outcomes of delinquency, depression, and alcohol use, in addition to providing genetic and nonshared environmental correlations between spanking and the these psychosocial outcomes. The results provide an original contribution to the spanking literature with regard to genetic and nonshared environmental covariation underpinning the associations between spanking and psychosocial outcomes, and these results serve as input parameters for simulation analyses in Study 2.

Method

Sample

Data for the present study are secured from the CNLSY sample. The CNLSY is a sample of youth born to a nationally representative sample of women from the National Longitudinal Survey of Youth 1979 (NLSY79). Children of mothers from the NLSY79 were assessed every two years from 1986 to 2014. Multiple children born to the same mother are included in the CNLSY. As such, many participants are biologically related to one another. Over the past decade, researchers have created a reliable and well-validated linking algorithm capable of identifying biologically related sibling pairs nested within the CNLSY. This algorithm has been used to assign levels of additive genetic relatedness between siblings in the CNLSY (Rodgers et al., 2016). The present study utilizes this algorithm to identify sibling pairs within the CNLSY in order to conduct a genetically informative analysis to assess the magnitude of additive genetic and environmental overlap between variation in spanking during childhood and delinquency, depression, and alcohol use in adolescence. The analytic sample consisted of 2868 respondents. Nested within these 2868 respondents are 934 full-sibling pairs and 500 half-sibling pairs. See Table SOM1 for descriptive statistics.

Measures

Spanking

Mothers in the CNLSY were asked during each assessment wave to report how often their one to 14-year old children were spanked in the past week, focusing on age intervals 4–5, 6–7, and 8–9 because this yielded the most complete longitudinal data on spanking for children in the CNLSY. Responses ranged from 0 to 18 for ages 4–9 with over 92% of responses ranging from 0 to 10. Values of 11–18 were recoded to have a value of *10* to create consistency in response categories across ages. Maternal-reported frequency of spanking during ages 4–5 was significantly and positively correlated with the frequency of spanking at ages 6–7 ($r = 0.37$, $p < 0.001$) and the frequency of spanking at ages 6–7 was significantly and positively correlated with frequency of spanking at ages 8–9 ($r = 0.29$, $p < 0.001$), albeit more weakly, suggesting the possibility of child-driven effects over time. Values across all age intervals were summed to create a cumulative measure of spanking for children between ages 4 and 9. To use data across time, the spanking measure was standardized and normalized at each age-interval using a Blom transformation, which employs a rank-based inverse normalizing transformation (Van den Oord et al., 2000). The measure used in the analysis represents the overall spanking score between ages 4 and 9 for each participant. Descriptive statistics for this measure before transformation are presented in Table SOM1.

Delinquency

From ages 14 to 17 years, participants were administered a self-report survey which asked them to report on their involvement in seven different types of delinquent behaviors during the past 12 months: (1) hurting someone on purpose; (2) lying to parents; (3) shoplifting; (4) destroying property at school; (5) skipping school; (6) staying out all night without permission, and (7) having a parent come to school because of their behavior. Response categories were no (0) or yes (1). Items were summed together to create variety indexes of delinquency at ages 14–15 and 16–17. Across ages 14–17, the median alpha for the index was 0.65. Both measures were summed to capture the frequency of self-reported delinquent behavior from ages 14–17. This measure has been used extensively by researchers (Connolly et al., 2015; Harden et al., 2012; Van Hulle et al., 2007) and validated by a previously published report showing that it significantly predicts criminal convictions among youth in the CNLSY sample (Lahey et al., 2006). Similar to the measure of spanking, the measure for delinquency from ages 14–17 was Blom-transformed to reduce nonnormality before behavior genetic analyses.

Depression

Depression was assessed at ages 14–15 and 16–17 by a 7-item version of the Center for Epidemiological Studies – Depression (CES-D) scale that asked participants how often in the past week they felt the following ways: (1) did not feel like eating;

(2) had trouble keeping mind on tasks; (3) felt depressed; (4) felt everything was an effort; (5) sleep was restless; (6) felt sad, and (7) could not "get going". Response categories for each item were 0 = *rarely, none of the time, 1 day,* 1 = *some, a little of the time, 1–2 days,* 2 = *occasionally, moderate amount of time, 3–4 days*, and 3 = *most, all of the time, 5–7 days*. Values were summed to reflect frequency of depressive symptoms at each age interval. The median alpha across ages 14–17 was 0.88. Scores for respondents at ages 14–15 and 16–17 were summed to create a measure of depressive symptomology from ages 14–17.

Alcohol use

Alcohol use was measured by asking participants to report, on average, how often in the past 12 months they consumed an alcoholic beverage including beer, wine, or liquor. Response categories were 0 = *did not drink alcohol in the past 12 months,* 1 = *1 to 2 days*, 2 = *3 to 5 days*, 3 = *every other month or so (6 to 11 days a year)*, 4 = *1 to 2 times a month (12 to 24 days a year)*, 5 = *several times a month (25 to 51 days a year)*, 6 = *1 or 2 days a week*, 7 = *3 to 6 days a week*, and 8 = *daily*. The average of alcohol use reported during ages 14–15 was 2.68 (3 to 5 days in the past month) and 3.41 (every month or so [6 to 11 days a year]) during ages 16–17. Values were summed to create a measure of average alcohol use from ages 14–17.

Demographic covariates

To control for confounding effects of family income, race, sex, and sibling-pair age difference, each variable was regressed on family income (total household income reported by mothers at age 30), race (0 = Black or Hispanic, 1 = Non-Black/Non-Hispanic), sibling-pair age difference (sibling 2 age [in years] – Sibling 1 age [in years]), and sex (0 = female, 1 = male). The standardized residuals from this regression-based analysis were used in subsequent univariate and bivariate quantitative behavioral genetic analyses. The racial composition of the sample was approximately 40% non-Black, non-Hispanic, 38% Black, and 21% Hispanic, and the sex composition was approximately 50% male and 50% female (see Table SOM1).

Analytic plan

The analysis for the present study was conducted in a series of three steps. First, phenotypic and within-sibling correlations for spanking, delinquency, depression, and alcohol use were examined. Phenotypic correlations were estimated with standard errors adjusted for non-independence since the data were secured on siblings from the same household. After inspecting bivariate associations between variables, within-sibling pair correlations were estimated. Within-sibling pair correlations assess the strength of association between siblings within a pair on a measurable phenotype. If within-sibling pair concordance on a phenotype is stronger for full siblings (who share, on average, 50% of their additive genetic material) than for half-siblings (who share, on average, 25% of their additive

genetic material), this can be interpreted as preliminary evidence that the phenotype under examination is under some degree of genetic influence.

Based on results from the phenotypic correlation and within-sibling pair analysis, univariate ACE models were then fitted to the data to partition the observed variation in spanking, delinquency, depression, and alcohol use into latent additive genetic, shared environment, and nonshared environment variance components. The latent additive genetic component (A) represents shared additive genetic variation between siblings, the shared environment component (C) represents family-level environmental experiences that make siblings similar to one another (e.g., culture, household family dynamics, socioeconomic status), and the nonshared environment component (E) represents individual-level environmental experiences unique to each sibling that create differences between siblings (e.g., different parent–child relationships, different peers, different school experiences). The nonshared environment variance component also includes the effects of systematic and random error. To estimate the magnitude of additive genetic influences, the correlation between the A components was fixed based on the amount of additive genetic material shared between siblings. The correlation between A components for full siblings was fixed to 0.50 and the correlation for half-siblings was fixed to 0.25. Because the shared environment reflects all shared experiences between siblings, the correlation between C components was fixed to 1.0; and the correlation between the E components was fixed to 0 because the nonshared environment assumes that siblings share 0% of their unique environment and these experiences are unique to each sibling. The magnitude of additive genetic, shared environment, and nonshared environment effects on an examined phenotype is calculated by comparing the observed cross-sibling correlations to the predicted cross-sibling correlations generated by the model. Model fit statistics are used to compare alternative biometric structures (AE and CE) to baseline ACE models.

The last step in the analysis involved fitting a series of bivariate Cholesky decomposition models to partition the covariance between spanking, delinquency, depression, and alcohol use into latent additive genetic, shared environment, and nonshared environment components. Figure SOM1 depicts a path diagram for the bivariate Cholesky decomposition model. Bivariate ACE models can provide information on the extent to which common and unique additive genetic and environmental variation explains the total amount of covariance between two phenotypes. The proportion of total variance in spanking due to genetic influences shared with delinquency, depression, and alcohol use was estimated by fitting a series of bivariate Cholesky decomposition models to the data.

All univariate and bivariate models were estimated using M*plus* 8.1 (Muthén & Muthén, 1998–2018) with full information maximum likelihood estimation and using the WLMSV estimator. Baseline models were compared alongside nested models using values from an adjusted χ^2 difference test (Santorra, 2000), Akaike Information Criteria (AIC), and Bayesian information criteria (BIC). A non-significant change in χ^2 indicates that the nested model with fewer parameters fits the data equally well and is the best-fitting model based on parsimony.

Table 1 Standardized portions of variance from univariate ACE models

	A	C	E	AIC	BIC	$\Delta\chi^2$	Δ *df*	*p*
Model								
Spanking								
ACE	**.36 (.08)**	**.26 (.07)**	**.38 (.03)**	**20,377.24**	**20,397.64**	**-**	**-**	**-**
AE	.49 (.10)	.00 (.00)	.51 (.03)	20,379.84	20,403.72	5.662	1	.026
CE	.00 (.00)	.32 (.12)	.68 (.03)	20,378.24	20,403.32	14.503	1	< .001
Delinquency								
ACE	.29 (.08)	.15 (.09)	.56 (.03)	20,618.11	20,720.79	-	-	-
AE	**.46 (.07)**	**.00 (.00)**	**.54 (.02)**	**20,618.09**	**20,707.88**	**5.893**	**1**	**.17**
CE	.00 (.00)	.10 (.08)	.90 (.02)	20,619.03	20,725.62	12.375	1	< .001
Depression								
ACE	.38 (.06)	.02 (.02)	.60 (.03)	22,437.57	22,540.65	-	-	-
AE	**.40 (.05)**	**.00 (.00)**	**.60 (.03)**	**22,434.86**	**22,525.05**	**5.608**	**1**	**.13**
CE	.00 (.00)	.05 (.04)	.95 (.04)	22,436.10	22,534.17	10.484	1	< .001
Alcohol use								
ACE	.21 (.07)	.08 (.06)	.71 (.02)	20,638.71	20,741.31	-	-	-
AE	**.26 (.05)**	**.00 (.00)**	**.74 (.02)**	**20,635.71**	**20,738.31**	**8.243**	**1**	**.25**
CE	.00 (.00)	.12 (.08)	.88 (.03)	20,637.37	20,739.15	10.675	1	< .001

Notes: Model estimator = maximum likelihood with robust standard errors. *AIC*, Akaike information criteria; *BIC*, Bayesian information criteria. *Df*, degrees of freedom; $\Delta\chi 2$, Santorra-Bentler scaled change in model chi-squared. Standardized parameter estimates presented. Standard errors in parentheses. Results control for race and sex. Best fitting model bolded

Results

Table SOM2 presents the phenotypic and intra-class correlations for full- and half-sibling pairs. Spanking during childhood was positively and moderately associated with delinquency ($r = 0.27$, 95% CI: 0.20–0.34), depression ($r = 0.17$, 95% CI: 0.11–0.28), and alcohol use ($r = 0.20$, 95% CI: 0.13–0.30) during adolescence. Full-siblings demonstrated stronger concordance on maternal reports of spanking ($r = 0.21$, 95% CI: 0.12–0.31), adolescent delinquency ($r = 0.33$, 95% CI: 0.26–0.42), adolescent depressive symptomology ($r = 0.30$, 95% CI: 0.25–0.37), and adolescent alcohol use ($r = 0.27$, 95% CI: 0.21–0.38) compared to half-siblings for spanking ($r = 0.16$, 95% CI: 0.10–0.20), adolescent delinquency ($r = 0.20$, 95% CI: 0.13–0.25), adolescent depressive symptomology ($r = 0.17$, 95% CI: 0.12–0.22), and adolescent alcohol use ($r = 0.19$, 95% CI: 0.10–0.25). Comparisons of same-sex and opposite-sex pairs for full- and half-siblings revealed overlapping confidence intervals suggesting non-significant differences in spanking, delinquency, depression, and alcohol use across same-sex and opposite-sex pairs. Results from univariate sex-limitation ACE models also revealed no evidence of significant sex differences in the magnitude of genetic influence on spanking, delinquency, depression, or alcohol use (results available upon request).

The next step of the analysis focused on decomposing the variance in all measures into latent additive genetic and environmental components. Table 1 presents

the results from all estimated univariate ACE models. The best-fitting model for spanking was a full ACE model, which suggested that 36% of the variation in spanking as due to latent additive genetic factors, 26% of the variation was due to shared environment, and 38% of the variance was due to nonshared environment (including error). Model fit statistics indicated that an AE model provided the best fit to the data for delinquency and suggested that 46% of the variation was due to additive genetic factors, while 54% of the variation was due to nonshared environment. An AE model also provided the best fit for depression suggesting that 40% of the variation in depressive symptomology was attributable to latent additive genetic factors, whereas 60% of the variation was attributable to nonshared environment factors. Lastly, variation in alcohol use was accounted for by genetic and nonshared environmental sources, such that 26% of the variation was attributable to latent additive genetic influences and 74% of the variation was attributable to nonshared environment.

The last step in the analysis focused on identifying genetic and environmental influences on spanking shared with delinquency, depression, and alcohol use. Standardized parameter estimates from all best-fitting bivariate models are presented in Table SOM3 and are used to calculate the genetic correlation (r_g), shared environment correlation (r_c), and nonshared environment correlation (r_e) as well as the proportion of covariance attributable to genes, shared environment, and nonshared environment. The r_g for spanking and delinquency was 0.54 (95% CI: 0.40–0.68), while the calculated r_e was 0.07 (95% CI: 0.03–0.14). There was no r_c between spanking and delinquency since constraining the c12 parameter to "0" improved model fit, whereby the AE model was the best fit to the data. As presented in Fig. 1, additive genetic covariation accounted for 66% of the phenotypic correlation between spanking and delinquency, and nonshared environment covariation accounted for the remaining 34% of the phenotypic association.

A r_g of 0.74 (95% CI: 0.59–0.90) and a r_e of 0.21 (95% CI: 0.10–0.41) was calculated for spanking and depression. Additive genetic covariation accounted for 48% of the phenotypic correlation between spanking and depression, and nonshared environment covariation accounted for 52% of the phenotypic correlation. With respect to spanking and alcohol use, the r_g and r_e estimates were 0.81 (95% CI: 0.62–0.92) and 0.05 (95% CI: 0.02–0.16), respectively. Additive genetic covariation accounted for 61% of the phenotypic correlation, and nonshared environment covariation accounted for 39% of the phenotypic correlation.

Study 2

Results of Study 1 indicate that the associations between spanking and psychosocial outcomes are substantially accounted for by both genetic and nonshared environmental covariation, consistent with previous studies using similar variables on different data sets (Button et al., 2008; Jaffee et al., 2004). A limitation of the genetically informed analyses on spanking and psychosocial outcomes is the limited range of outcomes assessed. The Gershoff and Grogan-Kaylor (2016a, b) meta-analysis makes clear, however, that the general spanking literature has

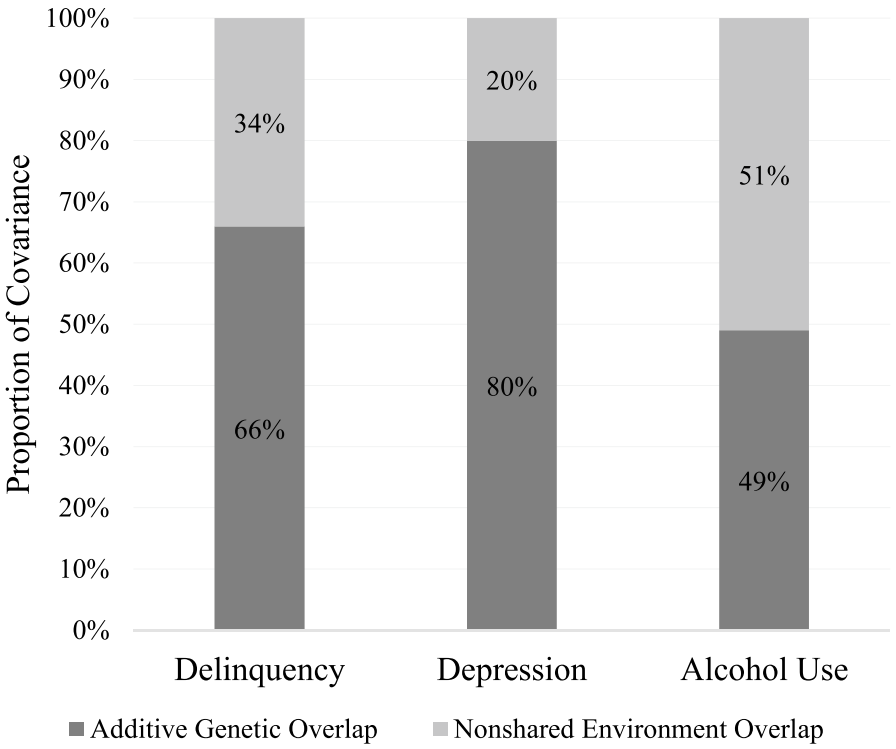

Fig. 1 Latent additive genetic and non-shared environment overlap between spanking and phenotypes. *Notes:* Proportions of covariance are calculated from coefficient estimates reported in Table SOM3. Estimates control for race and sex

investigated a broad range of outcomes showing direct associations with spanking. In an attempt to bridge the gap between the broader spanking literature and the behavior genetic spanking literature, Study 2 uses simulation modeling to contribute informed estimations of genetic and nonshared environmental covariation of the 18 phenotypic effects of spanking reported in the 2016 meta-analysis.

Study 2 uses a simulation tool and parameters available in the literature (i.e., heritability estimates, environmentality estimates, genetic correlations, nonshared environmental correlations, and phenotypic correlations) to estimate the degree to which the 18 phenotypic effects of spanking reported by Gershoff and Grogan-Kaylor (2016a, b) may be accounted for by genetic and nonshared environmental covariation. The output of the simulations contribute an evaluative summary of the current state of the literature on developmental outcomes of spanking from a genetically informed perspective, in complement to the recent, informative meta-analysis by Gershoff and Grogan-Kaylor. The primary contribution of this approach are summative estimates of genetic and nonshared environmental covariation on a much wider range of outcomes (see Gershoff & Grogan-Kaylor, 2016a, b) than has been available previously to researchers (Button et al., 2008;

Jaffee et al., 2004; Study 1), which will hopefully motivate continued research on genetic and nonshared environmental covariation on previously unexamined psychosocial outcomes.

Method

The *R* simulation tool developed by Barnes et al. (2017) allows calculations of the degree to which genetic covariation underpinning two traits is likely to explain an observed phenotypic effect. Such genetic covariation calculations have been used previously within the spanking literature (Button et al., 2008; Jaffee et al., 2004), and there are other similar tools available for these types of calculations as well (e.g., Purcell, 2016).

The present analysis uses the Barnes et al. (2017) simulation tool to estimate the degree to which the phenotypic association between spanking and the psychosocial outcomes reported in Gershoff and Grogan-Kaylor (2016a, b) can be accounted for by (1) shared genetic covariation, and (2) nonshared environmental covariation. We do not model the impact of shared environmental covariation given that previous analyses demonstrate very small proportions, if any of the phenotypic associations are accounted for by shared environmental covariation, with the overwhelming majority of the effect accounted for by genetic covariation and nonshared environmental covariation (Button et al., 2008; Jaffee et al., 2004).

The outputs of the simulations are (1) the estimated degree to which genetic covariation can account for the observed phenotypic effect, h^2_{cov}, and (2) the estimated degree to which nonshared environmental covariation can account for the observed phenotypic effect, e^2_{cov}. These outputs are the result of solving the equations below, with Eq. (1) assessing genetic covariation and Eq. (2) assessing nonshared environmental covariation. Each of these equations requires four pieces of information: (1) the phenotypic effect size (correlation) between the two traits, r_{p}, (2) the heritability estimate (Eq. (1)) or the nonshared environmental estimate (Eq. (2)) of the predictor variable, h^2_{x} or e^2_{x}, respectively, (3) the heritability estimate (Eq. (1)) or the nonshared environmental estimate (Eq. (2)) of the outcome variable, h^2_{y} or e^2_{y}, respectively, and (4) the genetic correlation (Eq. (1)) or the nonshared environmental correlation (Eq. (2)) between the two traits, r_{g} or r_{e}, respectively.

$$h^2_{cov} = \frac{\sqrt{h^2_x} * r_g * \sqrt{h^2_y}}{r_p} \tag{1}$$

$$e^2_{cov} = \frac{\sqrt{e^2_x} * r_e * \sqrt{e^2_y}}{r_p} \tag{2}$$

The input parameters can be obtained, to varying degrees, from the published literature.

First, we had to specify the observed phenotypic effects for the associations between spanking and psychosocial outcomes. For the present analysis, we relied on the meta-analytic effects reported in Gershoff and Grogan-Kaylor's (2016a, b). This meta-analysis provides standardized effect sizes for 17 specific psychosocial outcomes across the lifespan, and also provides an "overall effect" of spanking on psychosocial outcomes, more generally. Table 2 lists all 18 phenotypic effects reported by the Gershoff and Grogan-Kaylor meta-analysis. Of the phenotypic effects provided, 13 were demonstrated to be significant, meaning that exposure to spanking had a statistically significant adverse effect on psychosocial outcomes in childhood or adulthood. Phenotypic effect sizes (*d*) from Gershoff and Grogan-Kaylor are displayed in Table 2, with significant effects in bold.

We obtained heritability and nonshared environmental estimates for spanking behavior (the predictor variable), and for each of the 18 psychosocial outcome variables assessed by Gershoff and Grogan-Kaylor (2016a, b). For many psychosocial outcomes, we consulted the database of meta-analyzed heritability and environmentality estimates provided by Polderman et al. (2015). Of the 18 psychosocial variables examined in the present research, nine (50%) were included in the Polderman et al. database, and these estimates are provided in Table 2. Meta-analyzed heritability and environmentality estimates of six (33%) psychosocial outcomes were obtained from Burt (2009). The remaining heritability and environmentality estimates were obtained from individual published studies using academic search engines (e.g., Google Scholar, PsychInfo) to identify behavior genetic analyses of each psychosocial outcome. For the majority of the psychosocial outcomes examined, we identified multiple estimates from the published literature in addition to the estimates presented in meta-analyses. For the predictor variable of spanking, we obtained seven heritability and environmentality estimates from four measures across three separate studies (Barnes et al., 2013; Jaffee et al., 2004; Wade & Kendler, 2000). The specific measurement of each variable from each source is noted in the "Measure" column of Table 2.

Next, we converted the effect sizes from *d*, as reported in Gershoff and Grogan-Kaylor (2016a, b), to effect sizes of *r* using a standard *d*-to-*r* conversion equation (below), with the results listed in the *r* column of Table 2.

$$r = \sqrt{\frac{d^2}{4 + d^2}}$$

The final input parameter needed for the simulation tool is the genetic correlation, r_g, or nonshared environmental correlation, r_e, between each variable pair (i.e., spanking and a specific psychosocial outcome). Estimation of these correlations is limited in the literature. Because of this, Barnes et al. (2017) recommend using a range of plausible, yet conservative values. For the present research we implemented a range of $r_g = 0.30$–0.60 and $r_e = 0.00$–0.30 to be used across each psychosocial outcome based on results of Button et al. (2008), Jaffee et al. (2004), and estimates from Study 1.

Because the input parameters used for these calculations are never "fixed" given that heritability and environmental estimates, and phenotypic effects, can

Table 2 Summary of data used for simulation-modeling input

Outcome variable	*d*	*r*	h^2	*M* h^2	e^2	*M* e^2	Measure	Source
Low moral internalization	**.38**	.19	.31	.47	.42	.32	Meta-analysis—Social values	Polderman et al. (2015)
			.37, .26, .30, .47, .51, .51, .52, .60, .62. .72,		.20, .30, .40, .31, .39, .28, .24, .32, .31, .38		Revised Rutter Parent Scale for Preschool Children—Prosocial Behavior (ages 2, 3, 4, 7)	Knafo and Plomin (2006)
Child aggression	**.37**	.18	.47, .48, .44, .40	.49	.25, .28, .27, .25	.26	Devereux Child Behavior rating scale (Aggressive behavior subscale)	van Beijsterveldt et al. (2004)
Child antisocial behavior	**.39**	.19	.73	.66	.27	.28	Achenback family of instruments and DSM-V	Jaffee et al. (2004)
			.58		.28		Meta-analysis—Conduct problems	Burt (2009)
Child externalizing behavior problems	**.41**	.20	.70, .78	.63	.30, .22	.22	Preschool & Kindergarten Behavior Scale (2ed; 7-itmes)—Parent reports	Barnes et al. (2013)
			.42, .82		.27, .18		Preschool and Kindergarten Behavior Scale (2ed; 6-itmes)—Teacher reports	
			.52, .56		.16, .14		The Child Behavior Checklist (external problem items only)	Van der Valk et al. (2003)
			.59		.26		Meta-analysis—Externalizing	Burt (2009)
Child internalizing behavior problems	**.24**	.12	.35, .38	.41	.30, .32	.32	The Child Behavior Checklist (internal problem items only)	Van der Valk et al. (2003)
			.51		.33		Meta-analysis—Internalizing	Burt (2009)

Table 2 (continued)

Child mental health problems	**.53**	.26	.20, .23	.36	.53, .57	.45	Center for Epidemiological Studies Depression Scale and Zung Anxiety Scale (10–12 years)	Zheng et al. (2016)
			.44		.42		Meta-analysis—Depression	Burt (2009)
			.48		.40		Meta-analysis—Anxiety	Burt (2009)
			.44		.36		Meta-analysis—Emotional disorders with childhood onset	Polderman et al. (2015)
Negative parent–child relationship	**.51**	.25	.19	.25	.62	.50	Meta-analysis—Family relationships	Polderman et al. (2015)
			.30		.37		Meta-analysis—Problems related to upbringing	
Impaired cognitive ability	**.17**	.08	.51	.51	.25	.25	Meta-analysis—Higher-order mental functions	Polderman et al. (2015)
Low self-esteem	**.15**	.07	.36, .40	.37	.60, .64	.62	Rosenberg Self-Esteem Inventory	Roy et al. (1995)
			.29, .32		.66, .71			Kendler et al. (1998)
			.47		.51			Neiss et al. (2006)
Victim of physical abuse	**.64**	.30	.07	.07	.05	.05	Parent interview	Jaffee et al. (2004)
Adult antisocial behavior	**.36**	.18	.47	.48	.36	.35	Meta-analysis—Mental and behavioral disorders	Polderman et al. (2015)
			.49		.33		Meta-analysis—Conduct disorders	
Adult mental health problems	**.24**	.12	.47	.49	.36	.41	Meta-analysis—Mental and behavioral disorders	Polderman et al. (2015)
			.61		.31		Meta-analysis—Mood disorders	
			.38		.55		Meta-analysis—Dissocial personality disorder	

Table 2 (continued)

Adult support for physical punishment	**.38**	.19	.31	.31	.50	.50	Meta-analysis—Attitudes	Polderman et al. (2015)
Overall effect size	**.33**	.16	.45	.45	.40	.40	Meta-Analysis—Mental functions (Big-5 Personality and IQ)	Polderman et al. (2015)
Predictor variable			h^2	$M\ h^2$	e^2	$M\ e^2$	Measure	Source
Spanking			.25	.31	.10	.19	Parent interview	Jaffee et al. (2004)
			.37		.39		Child reports of parent spanking	Wade et al. (2000)
			.16		.21		Parent reports of spanking	
			.27, .50		.07, .18		Parent report of spanking frequency	Barnes et al. (2013)

fluctuate across time, space, and sample, the simulation tool (Barnes et al., 2017) allows for constructing beta-distributions for the input values of r_p, h^2_x,, h^2_y, e^2_x, and e^2_y. (Barnes et al. provide formal mathematical details for building the beta-distributions to use the simulation tool.) The beta-distributions allow for natural range and unspecified error (including measurement error, sampling bias) to be taken into account when calculating h^2_{cov} and e^2_{cov}. The beta-distributions for r_ps are approximately centered on the meta-analyzed effect size reported by Gershoff and Grogan-Kaylor (2016a, b). The beta-distributions for the heritability and environmentality inputs, which account for variation of estimates around the simple mean, are approximately centered on the simple arithmetic means for the heritability and environmentality estimates of each variable (provided in the M h^2 and M e^2 column of Table 2). The beta-distributions for the overall effect of spanking are shown in Figure SOM2 and SOM3, and beta-distributions for the 17 remaining specific outcomes are provided and in Figure SOM4a-m and Figure SOM5a-m.

When solving for h^2_{cov} or e^2_{cov} at a specified value of r_g or r_e, respectively, 10,000 calculations are computed. Each calculation draws a random value of r_p, h^2_x, and h^2_y, from the respective beta-distributions for the genetic covariation simulation, and a random value of r_p, e^2_x, and e^2_y, from the respective beta-distributions for the non-shared environmental covariation simulation. Put differently, 10,000 calculations are computed at each value of r_g at intervals of 0.01 beginning with $r_g = 0.30$ through $r_g = 0.60$, and at each value of r_e at intervals of 0.01 beginning with $r_e = 0.00$ through $r_e = 0.30$. Therefore, in the present analysis, 310 thousand calculations are computed for each phenotypic effect, resulting in more than 11 million calculations in total with this tool. The output, described below, is the mean value and 95% credible interval of the 10,000 h^2_{cov} and e^2_{cov} estimates at each value of r_g or r_e, respectively.

Results

Shared genetic covariation

The simulation output collectively suggests that the phenotypic effects of spanking on psychosocial outcomes are likely to be explained substantially by shared genetic covariation underpinning both traits (see Figure SOM6a-k for figures corresponding to each outcome variable). For all psychosocial outcomes except one (discussed below), average estimates of genetic covariation were greater than 50%, given the specified phenotypic effect (see Table 2 across the range of genetic correlations indicated by previous research and the results of Study 1, $r_g = 0.30$–0.60). Examining the model output for the overall effect of spanking, it can be observed that at the given effect size ($d = 0.33$, $r_p = 0.16$), nearly all of the phenotypic covariation will likely be accounted for by shared genetic covariation between spanking and the trait of interest once genetic correlations become large in size, $r_g > 0.50$ (see Fig. 2). There were traits for which the simulation output suggested that genetic covariation could potentially explain the entire effect — meaning that the mean h^2_{cov} and associated

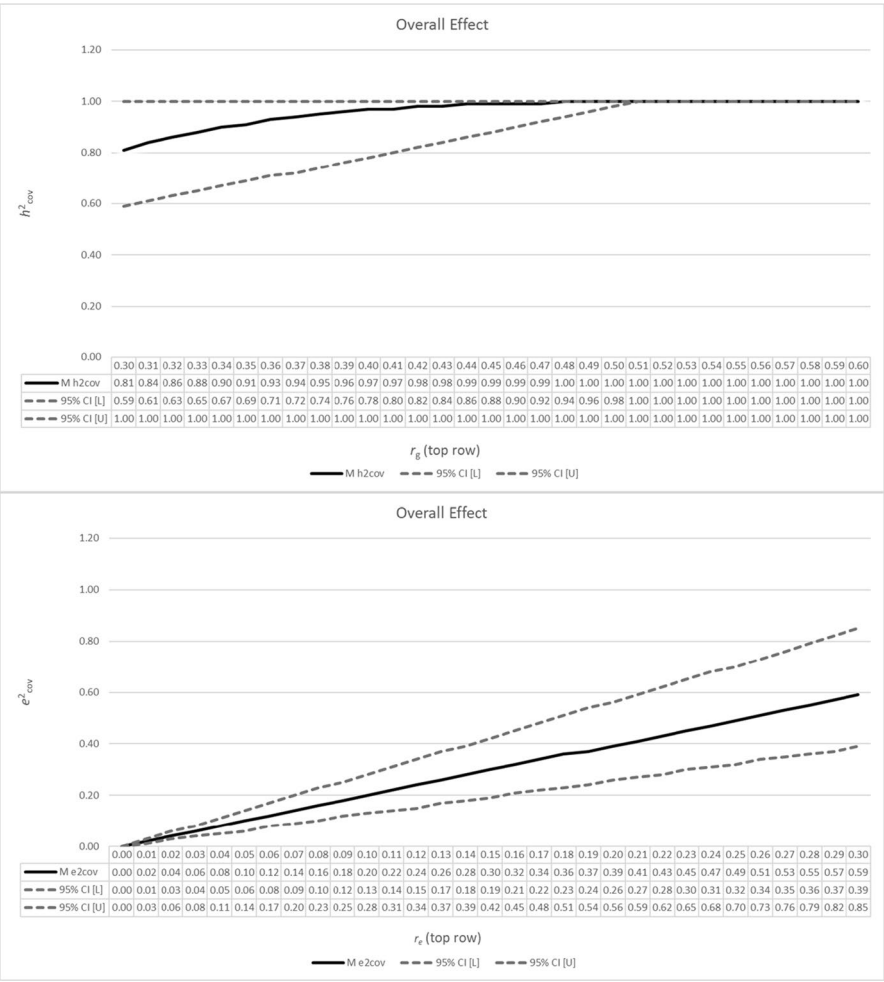

Fig. 2 Top: Mean h^2_{cov} (black, solid line) with 95% credibility intervals (gray, dashed lines) for overall effect of spanking on child and adult outcomes at different values of r_g assuming effect size of $d = .33$, $r_p = .16$. Bottom: Mean e^2_{cov} (black, solid line) with 95% credibility intervals (gray, dashed lines) for overall effect of spanking on child and adult outcomes at different values of r_e assuming effect size of $d = .33$, $r_p = .16$

95% credible interval were all equal to 1. These traits, impaired cognitive ability and self-esteem, had very small phenotypic effect sizes, $r_p < 0.1$.

The one exception to the overall trend of substantial genetic covariation explaining the phenotypic effect was the outcome victim of physical abuse. This outcome showed, overall, low explanation by genetic covariation, unlike the rest of the variables assessed, such that even at the high end of the genetic correlation range ($r_g = 0.60$), genetic covariation was estimated at only approximately 20%.

Non-shared environmental covariation

The simulation output collectively suggests that the phenotypic effects of spanking on psychosocial outcomes are likely to be explained by small-to-moderate degrees of nonshared environmental covariation underpinning both traits of interest (see Figure SOM7a-m for figures corresponding to each outcome variable). For most significant psychosocial outcomes except one (discussed below), average estimates of nonshared environmental covariation ranged from 0 to 40%, with some outcomes showing nonshared environmental covariation of 80% or greater, given the specified phenotypic effect (see Table 2) across the range of nonshared environmental correlations indicated by previous research and the results of Study 1, $r_e = 0.00–0.30$. Examining the model output for the overall effect of spanking, in particular, it can be observed that at the given effect size ($d = 0.33$, $r_p = 0.16$), nonshared environmental covariation may explain small-to-moderate proportions of the phenotypic effect between spanking and psychosocial outcomes. The simulation output yielded very high estimates of nonshared environmental covariation for self-esteem, which had a very small phenotypic effect size in the Gershoff and Grogan-Kaylor (2016a, b) meta-analysis — meaning that the mean e^2_{cov} and associated upper-bound 95% credible interval were equal to 1 at the upper end of the r_e range assessed here (approaching $r_e = 0.30$; see Fig. 2).

The one exception to the overall trend of moderate nonshared environmental covariation explaining the phenotypic effect was, again, the outcome victim of physical abuse. This outcome showed, overall, low explanation by nonshared environmental covariation, unlike the rest of the variables assessed, such that even at the high end of the nonshared environmental correlation range ($r_e = 0.30$), nonshared environmental covariation was estimated at only approximately 10%.

General discussion

The current research analyzed sibling data from the Children of the National Longitudinal Survey of Youth (CNLSY; Study 1) and conducted simulation models using input parameters from the existing literature (Study 2) to provide a summative evaluation of genetic and nonshared environmental covariation explaining the associations between spanking and psychosocial outcomes. Study 1 found moderate effects of spanking on outcomes of adolescent delinquency, depression, and alcohol use, with results suggesting a moderate-to-large degree of genetic covariation accounting for the associations, and a small-to-moderate degree of nonshared environmental covariation accounting for the associations. Study 2 estimated plausible ranges of genetic and nonshared environmental covariation on a broad range of psychosocial outcomes of spanking reported in a meta-analysis by Gershoff and Grogan-Kaylor (2016a, b). Broadly, the simulations suggest that the effects of spanking on psychosocial outcomes can, in several instances, be explained largely by genetic covariation, with the remainder attributable to nonshared environmental covariation.

The results of the analyses on the CNLSY data and the simulations are consistent with results by Jaffee et al. (2004) and Button et al. We find moderate effects of

spanking on delinquent behavior, such that the majority of the phenotypic effect was attributable to genetic covariation, and the remainder attributable to nonshared environmental covariation. The simulation models using the most up-to-date effect sizes of the effects of spanking on a wide range of outcomes provide a more comprehensive understanding of the negative effects of spanking. In the context of the broader behavior genetic literature on spanking, it can reasonably be concluded that genetic covariation explains a non-trivial proportion of the covariance between spanking and psychosocial outcomes, with the remaining proportion accounted for by nonshared environmental covariation.

This pattern of results is largely due to larger observed genetic correlations between spanking and psychosocial outcomes as compared to the smaller observed nonshared environmental correlations between the traits (see Study 1; see also, Button et al., 2008; Jaffee et al., 2004). The overall findings produce two key points. First, it is both plausible and probable that spanking exerts a negative causal impact on various developmental outcomes across the life course. Like most prior studies, moreover, we find no evidence that physical punishment *improves* child outcomes. Second, and from a methodological perspective, continued investigation into the developmental outcomes of spanking is important, but it is imperative for work in this area to use sibling-based research designs more frequently. Doing so will increase the precision of effect size estimation as it relates to spanking and child outcomes.

An exception to this general conclusion is the issue of spanking effects versus those of physical abuse, a point raised early on and to which we now return. Behavioral genetic research has demonstrated compellingly that (1) spanking and physical abuse show non-trivial shared environmental effects (Jaffee et al., 2004; Study 1), (2) physical abuse has robust, negative effects on psychosocial outcomes that appear to be causal (Lynch et al., 2006), and relatedly (3) the association between spanking and physical abuse is primarily explained by *shared* environmental covariation (Jaffee et al., 2004). Our simulation output yielded small estimates of genetic ($\approx$ 20%) and nonshared environmental covariation ($\approx$ 10%), suggesting that *shared* environmental covariation indeed largely explains the phenotypic effect between spanking and physical abuse, consistent with previous work (Jaffee et al., 2004; Lynch et al., 2006).

Research implications and future directions

The evidence of genetic covariation explaining a substantial proportion of the associations between spanking and psychosocial outcomes (with the exception of physical abuse) has implications for the magnitude of the reported effects of spanking on development. Overall, the pattern of results suggesting moderate-to-large degrees of genetic covariation implies that once genetic overlap is considered, the phenotypic effects of spanking on certain psychosocial outcomes will be smaller in magnitude than what non-genetically informed studies have suggested (see also, Lynch et al., 2006). Specifically, meta-analyzed effect sizes of spanking on psychosocial outcomes from studies using familial designs will likely be smaller in magnitude than

those reported in Gershoff and Grogan-Kaylor (2016a, b). This point, while important, should not be misunderstood. We expand on this issue more fully when concluding, however, the existing evidence, as well as the data reported here, strongly suggest that nonshared environmental covariation explains a substantive proportion of the phenotypic effects between spanking and developmental outcomes—a finding that is consistent with the possibility that spanking (in addition to abuse) and certain developmental outcomes are *causally linked* and not merely *associated* with each other (see Turkheimer & Harden, 2013).

That said, our simulation models highlight the minimal overlap between the spanking literature more broadly, and the behavior genetic literature on the subject, specifically. To date, a considerably more limited range of psychosocial outcomes have been assessed using twin and sibling data (Button et al., 2008; Cecil et al., 2012; Jaffee et al., 2004; Lynch et al., 2006; O'Connor et al., 1998; Riggins-Casperset et al., 2003) in comparison with the wider array of outcomes examined utilizing a standard social science approach (Barnes et al., 2014a; Gershoff & Grogan-Kaylor, 2016a, b; Lee, 2012). To illustrate the point, consider that outcomes such as internalizing problems, negative parent–child relationship, cognitive impairments, self-esteem, and adult support for physical punishment, as far as we are currently aware, have not yet been subjected to analyses capable of fully controlling for heritable confounding.

The output of our simulations for these yet unstudied variables is consistent with the pattern of findings observed for the more commonly examined variables of externalizing behaviors. Given the consistency in the pattern of results, it seems reasonable to predict that genetic covariation will account for some non-zero proportion of the phenotypic effect of spanking on a wider range of measures. Indeed, the primary contribution of the simulation models is to demonstrate the plausibility and estimated magnitude of genetic covariation across a broader range of psychosocial outcomes, and to further highlight the usefulness of sibling-based designs for research on physical punishment.

As the genetically informed literature in this area expands, a key next step will be an updated formal meta-analysis of effect sizes for studies that have more fully accounted for familial confounding. The non-genetically informative literature does show wide ranging negative effects of spanking on psychosocial outcomes (Gershoff & Grogan-Kaylor, 2016a, b), yet the behavior genetic literature is not yet mature enough to evaluate alongside the existing meta-analysis. Here, we have contributed additional behavior genetic data (see Study 1) in an effort to add to the dearth of available studies. Our goal with the inclusion of the simulation analyses in Study 2 is largely to demonstrate the need for continued genetically sensitive research on spanking and child development outcomes.

A final consideration to note is that the spanking literature would benefit from more rigorous longitudinal designs, which would aid in the task of investigating evocative child effects (e.g., Cecil et al., 2012). Much of the literature has been built on cross-sectional and retrospective studies (Gershoff & Grogan-Kaylor, 2016a, b). Available longitudinal work (Barnes et al., 2013; Cecil et al., 2012) and adoption studies (Hao & Matsueda, 2006; O'Connor et al., 1998; Riggins-Casperset et al., 2003) seem to suggest that a bi-directional association is likely to exist between

spanking and psychosocial outcomes. Although the age at which spanking is received does not appear to have a large impact on the magnitude of associations between spanking and psychosocial outcomes (Gershoff & Grogan-Kaylor, 2016a, b), evidence from Cecil et al. (2012) also seems to indicate that enduring effects of spanking may be smaller relative to effects in childhood—however, longitudinal studies are needed to better clarify the role of age and enduring effects on a wide range of outcomes. Ideally, longitudinal behavior genetic studies assessing a wide range of outcomes would form the path forward for research in this area in a way that truly deepens causal inference abilities and offers more clarity on the complexities of punishment use and child development (see also, Gershoff et al., 2018).

Limitations

The primary limitation of the current research is the boundary conditions within which the simulation output must be interpreted. As with all simulations, the output relies on the accuracy of the inputs. One way in which we worked to be as inclusive as possible with regard to variation of input parameters was to build and sample from beta-distributions, rather than using only a single input value for each trait. The simulation output is calculated from thousands of random values drawn from each distribution, which allow for natural variation in each value due to error, time, space, and sample (Barnes et al., 2017). The beta-distributions themselves were also largely built from values obtained from meta-analyses (Burt, 2009; Polderman et al., 2015), thus providing the most accurate values currently available in the literature.

The most important note with regard to interpreting the simulation output is that the output is most informative within the range of the phenotypic effect (r_p = association between spanking and outcome) specified in the beta-distribution, which is centered on the meta-analyzed effect sizes from Gershoff and Grogan-Kaylor (2016a, b). Studies yielding larger phenotypic effect sizes will yield smaller genetic covariation at the same r_g values used here. Larger phenotypic effect sizes require correspondingly larger r_g to explain similar proportions of the phenotypic effect sizes (Barnes et al., 2014a).

Our interpretations of the extent to which the phenotypic effects may be explained by shared genetic covariation assume that the true genetic correlation falls within the range specified in our models, which is 0.30–0.60. This range was chosen based on the only three available studies for which genetic correlations have been calculated between spanking and some of the psychosocial outcomes assessed by Gershoff and Grogan-Kaylor (including Study 1, here). If the true genetic correlation between any two phenotypes (e.g., spanking and moral internalization) is smaller than what we have specified, then the extent to which genetic covariation can explain the phenotypic effect will be reduced. The extent to which the phenotypic effects reported in Gershoff and Grogan-Kaylor would be explained by genetic covariation if the genetic correlation were assumed to be smaller, r_g = 0.00–0.25, are provided in Table SOM3a-f.

With regard to Study 1, a limitation is the use of full siblings and half siblings, rather than a more complete quantitative genetic design employing MZ and DZ twins. On the one hand, we were limited by the nature of the dataset, as only a handful of twin pairs with available data on the variables of interest were included in the sample. Nonetheless, this characteristic of the data introduces the possibility of bias in the calculation of our genetic correlations. The two primary sources of bias in quantitative genetic research involve failure to satisfy the equal environments assumption (EEA; Bhattacharjee & Sarkar, 2017) and the assumption of random mating (for the phenotypes being examined) (Barnes et al., 2014a, 2014b). For humans, an assumption of random mating seems warranted rarely, if ever, and thus results in the possibility of obtaining downwardly biased heritability coefficients. The EEA, when violated, functions to artificially inflate heritability coefficients.

As a recent in-depth analysis suggested, however, even intentional violations of *both* the EEA and random mating assumptions that resulted in relatively minimal bias in the parameter estimates obtained using twin designs (Barnes et al., 2014a, 2014b). Even *if* the nature of analyzing full siblings and half siblings resulted in a higher likelihood of violating one, or both, assumptions, the results would remain substantively unchanged. As some final points of consideration, it is important to recall that parameter estimates in sibling studies can and do fluctuate depending on historical context and other cultural factors—they are neither fixed nor static values (Barnes et al., 2014b; Tucker-Drob & Bates, 2016; Turkheimer, 2000). Heritability estimates for cognitive traits, for example, can vary depending on broader socioeconomic factors to which children are exposed (see Tucker-Drob & Bates, 2016). Given this, it seems plausible that additional complexities will arise when examining the intersections of heritability, spanking, and location in various socioeconomic strata, an issue left unexamined in our study. The general point is that our results are informed by the culture and time period in which the data were collected, and are subject to vary when the consideration of additional factors is undertaken.

Conclusions

A voluminous body of work has documented robust associations between spanking and various negative psychosocial outcomes (Gershoff & Grogan-Kaylor, 2016a, b). A comparatively limited subset of this work, however, has used designs capable of fully accounting for sources of familial confounding, both heritable and environmental (Button et al., 2008; Cecil et al., 2012; Jaffee et al., 2004; Lynch et al., 2006; O'Connor et al., 1998; Riggins-Caspers et al., 2003). The results of Study 1 replicate previous work (e.g., Jaffee et al., 2004) revealing associations between spanking and delinquency, depression, and alcohol use, and finds that these associations are partly explained by moderate degrees of genetic covariation and small-to-moderate degrees of nonshared environmental covariation. Study 2 used a simulation modeling procedure (Barnes et al., 2017) to provide estimates of genetic and nonshared environmental covariation underpinning recently meta-analyzed phenotypic effects of spanking across a plausible range of genetic and nonshared environmental correlations.

The simulation output estimates suggested that, in most cases, genetic covariation likely accounted for a substantive portion of the observed phenotypic effect between spanking and the outcome of interest (between 60 and 80%), with the remainder attributable to nonshared environmental covariation (upwards of 40%). The findings of the current research indicate that continued investigation of the consequences of spankings would be best served by taking advantage of behavior genetic designs to assess a wide range of outcomes so as to continue clarifying the complex associations between spanking and development. That said, one can presently observe at least two overarching points, one methodological and one translational. Based on our work here, and that of others, it may be that spanking exerts a more *limited* range of causal effects on developmental outcomes than previously anticipated. Yet, the existing body of work stands as sufficient to undermine any lingering presumptions that spanking constitutes a method for promoting *healthy* development and child wellness.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11292-021-09496-5.

References

Afifi, T. O., Ford, D., Gershoff, E. T., Merrick, M., Grogan-Kaylor, A., Ports, K. A., MacMillan, H. L., Holden, G. W., Taylor, C. A., Lee, S. J., & Peters Bennett, R. (2017). Spanking and adult mental health impairment: The case for the designation of spanking as an adverse childhood experience. *Child Abuse and Neglect, 71*, 24–31.

Barnes, J. C., Barbaro, N., Boutwell, B. B., & Shackeford, T. K. (2017). Technical appendix for: Genetic confounding of the relationship between father absence and age at menarche. *Evolution and Human Behavior, 38*,. https://doi.org/10.1016/j.evolhumbehav.2016.11.007

Barnes, J. C., Boutwell, B. B., Beaver, K. M., & Gibson, C. L. (2013). Analyzing the origins of childhood externalizing behavioral problems. *Developmental Psychology, 49*, 2272–2284.

Barnes, J. C., Boutwell, B. B., Beaver, K. M., Gibson, C. L., & Wright, J. P. (2014a). On the consequences of ignoring genetic influences in criminological research. *Journal of Criminal Justice, 42*, 471–482.

Barnes, J. C., Wright, J. P., Boutwell, B. B., Schwartz, J. A., Connolly, E. J., Nedelec, J. L., & Beaver, K. M. (2014b). Demonstrating the validity of twin research in criminology. *Criminology, 52*, 588–626.

Benjet, C., & Kazdin, A. E. (2003). Spanking children: The controversies, findings, and new directions. *Clinical Psychology Review, 23*, 197–224.

Bhattacharjee S. & Sarkar A. (2017). Equal environments assumption. In: Vonk J., Shackelford T. (Eds.), *Encyclopedia of Animal Cognition and Behavior*. Springer.

Boutwell, B. B., & Adams, C. D. (2020). A research note on Mendelian randomization and causal inference in criminology: Promises and considerations. *Journal of Experimental Criminology*, 1–12.

Burt, S. A. (2009). Rethinking environmental contributions to child and adolescent psychopathology: A meta-analysis of shared environmental influences. *Psychological Bulletin, 135*, 608–637.

Button, T. M., Lau, J. Y., Maughan, B., & Eley, T. C. (2008). Parental punitive discipline, negative life events and gene–environment interplay in the development of externalizing behavior. *Psychological Medicine, 38*, 29–39.

Cecil, C. A., Barker, E. D., Jaffee, S. R., & Viding, E. (2012). Association between maladaptive parenting and child self-control over time: Cross-lagged study using a monozygotic twin difference design. *The British Journal of Psychiatry, 201*, 291–297.

Chabris, C. F., Lee, J. J., Cesarini, D., Benjamin, D. J., & Laibson, D. I. (2015). The fourth law of behavior genetics. *Current Directions in Psychological Science, 24*, 304–312.

Connolly, E. J., Schwartz, J. A., Nedelec, J. L., Beaver, K. M., & Barnes, J. C. (2015). Different slopes for different folks: Genetic influences on growth in delinquent peer association and delinquency during adolescence. *Journal of Youth and Adolescence, 44*, 1413–1427.

Danese, A., & Tan, M. (2014). Childhood maltreatment and obesity: Systematic review and meta-analysis. *Molecular Psychiatry, 19*(5), 544–554.

Durrant, J., & Ensom, R. (2012). Physical punishment of children: Lessons from 20 years of research. *Canadian Medical Association Journal, 184*, 1373–1377.

Ferguson, C. J. (2013). Spanking, corporal punishment and negative long-term outcomes: A meta-analytic review of longitudinal studies. *Clinical Psychology Review, 33*, 196–208.

Gershoff, E. T. (2002). Corporal punishment by parents and associated child behaviors and experiences: A meta-analytic and theoretical review. *Psychological Bulletin, 128*, 539–579.

Gershoff, E. T. (2013). Spanking and child development: We know enough now to stop hitting our children. *Child Development Perspectives, 7*, 133–137.

Gershoff, E. T., Goodman, G. S., Miller-Perrin, C. L., Holden, G. W., Jackson, Y., & Kazdin, A. E. (2018). The strength of the causal evidence against physical punishment of children and its implications for parents, psychologists, and policymakers. *American Psychologist, 73*, 626–638.

Gershoff, E. T., & Grogan-Kaylor, A. (2016a). Race as a moderator of associations between spanking and child outcomes. *Family Relations, 65*, 490–501.

Gershoff, E. T., & Grogan-Kaylor, A. (2016b). Spanking and child outcomes: Old controversies and new meta-analyses. *Journal of Family Psychology, 30*, 453.

Grogan-Kaylor, A., Ma, J., & Graham-Bermann, S. A. (2018). The case against physical punishment. *Current Opinion in Psychology, 19*, 22–27.

Hao, L., & Matsueda, R. L. (2006). Family dynamics through childhood: A sibling model of behavior problems. *Social Science Research, 35*, 500–524.

Harden, K. P., Quinn, P. D., & Tucker-Drob, E. M. (2012). Genetically influenced change in sensation seeking drives the rise of delinquent behavior during adolescence. *Developmental Science, 15*, 150–163.

Holland, W. (1986). Statistics and causal inference. *Journal of the American Statistical Association, 81*, 945–960.

Infurna, M. R., Reichl, C., Parzer, P., Schimmenti, A., Bifulco, A., & Kaess, M. (2016). Associations between depression and specific childhood experiences of abuse and neglect: A meta-analysis. *Journal of Affective Disorders, 190*, 47–55.

Jaffee, S. R., Caspi, A., Moffitt, T. E., Polo-Tomas, M., Price, T. S., & Taylor, A. (2004). The limits of child effects: Evidence for genetically mediated child effects on corporal punishment but not on physical maltreatment. *Developmental Psychology, 40*, 1047.

Kendler, K. S., Gardner, C. O., & Prescott, C. A. (1998). A population-based twin study of self-esteem and gender. *Psychological Medicine, 28*, 1403–1409.

Knafo, A., & Plomin, R. (2006). Prosocial behavior from early to middle childhood: Genetic and environmental influences on stability and change. *Developmental Psychology, 42*, 771.

Lahey, B. B., Van Hulle, C. A., Waldman, I. D., Rodgers, J. L., D'Onofrio, B. M., Pedlow, S., & Keenan, K. (2006). Testing descriptive hypotheses regarding sex differences in the development of conduct problems and delinquency. *Journal of Abnormal Child Psychology, 34*, 730–748.

Lansford, J. E., Wager, L. B., Bates, J. E., Pettit, G. S., & Dodge, K. A. (2012). Forms of spanking and children's externalizing behaviors. *Family Relations, 61*, 224–236.

Larzelere, R. E., & Kuhn, B. R. (2005). Comparing child outcomes of physical punishment and alternative disciplinary tactics: A metaanalysis. *Clinical Child and Family Psychology Review, 8*, 1–37.

Larzelere, R. E., Kuhn, B. R., & Johnson, B. (2004). The intervention selection bias: An underrecognized confound in intervention research. *Psychological Bulletin, 130*, 289–303.

Lee, J. J. (2012). Correlation and causation in the study of personality. *European Journal of Personality, 26*(4), 372–390.

Lynch, S. K., Turkheimer, E., D'onofrio, B. M., Mendle, J., Emery, R. E., Slutske, W. S., & Martin, N. G. (2006). A genetically informed study of the association between harsh punishment and offspring behavioral problems. *Journal of Family Psychology, 20*, 190–198.

Muthén, L. K., & Muthén, B. O. (1998–2018). *Mplus user's guide.* Los Angeles, CA: Muthén and Muthén.

Neiss, M. B., Sedikides, C., & Stevenson, J. (2006). Genetic influences on level and stability of self-esteem. *Self and Identity, 5*, 247–266.

O'Connor, T. G., Deater-Deckard, K., Fulker, D., Rutter, M., & Plomin, R. (1998). Genotype–environment correlations in late childhood and early adolescence: Antisocial behavioral problems and coercive parenting. *Developmental Psychology, 34*, 970–981.

Paolucci, E. O., & Violato, C. (2004). A meta-analysis of the published research on the affective, cognitive, and behavioral effects of corporal punishment. *The Journal of Psychology, 138*, 197–221.
Plomin, R., DeFries, J. C., Knopik, V. S., & Neiderhiser, J. M. (2016). Top 10 replicated findings from behavioral genetics. *Perspectives on Psychological Science, 11*, 3–23.
Plomin, R., DeFries, J. C., Knopik, V. S., & Neiderhiser, J. M. (2013). *Behavioral genetics* (6th ed.). Worth Publishers.
Polderman, T. J. C., Benyamin, B., de Leeuw, C. A., Sullivan, P. F., van Bochoven, A., Visscher, P. M., & Posthuma, D. (2015). Meta-analysis of the heritability of human traits based on fifty years of twin studies. *Nature Genetics, 47*, 702–709.
Purcell, S. (2016). Statistical methods in behavioral genetics. In R. Plomin, J.C. DeFries, V.S. Knopik, & J.M. Neiderhiser, Behavioral Genetics, Sixth Edition (pp. 357–412).
Riggins-Caspers, K. M., Cadoret, R. J., Knutson, J. F., & Langbehn, D. (2003). Biology-environment interaction and evocative biology-environment correlation: Contributions of harsh discipline and parental psychopathology to problem adolescent behaviors. *Behavior Genetics, 33*, 205–220.
Rodgers, J. L., Beasley, W. H., Bard, D. E., Meredith, K. M., Hunter, M. D., Johnson, A. B., & Miller, W. B. (2016). The NLSY kinship links: Using the NLSY79 and NLSY-children data to conduct genetically-informed and family-oriented research. *Behavior Genetics, 46*, 538–551.
Rohrer, J. M. (2018). Thinking clearly about correlations and causation: Graphical causal models for observational data. *Advances in Methods and Practices in Psychological Science, 1*(1), 27–42.
Roy, M. A., Neale, M. C., & Kendler, K. S. (1995). The genetic epidemiology of self-esteem. *The British Journal of Psychiatry, 166*, 813–820.
Scarr, S., & McCartney, K. (1983). How people make their own environments: A theory of genotype environment effects. *Child Development, 54*, 424–435.
Tucker-Drob, E. M., & Bates, T. C. (2016). Large cross-national differences in gene× socioeconomic status interaction on intelligence. *Psychological Science, 27*(2), 138–149.
Turkheimer, E. (2000). Three laws of behavior genetics and what they mean. *Current Directions in Psychological Science, 9*, 160–164.
Van Beijsterveldt, C. E. M., Verhulst, F. C., Molenaar, P. C. M., & Boomsma, D. I. (2004). The genetic basis of problem behavior in 5-year-old Dutch twin pairs. *Behavior Genetics, 34*, 229–242.
Van der Valk, J. C., Van den Oord, E. J. C. G., Verhulst, F. C., & Boomsma, D. I. (2003). Using shared and unique parental views to study the etiology of 7-year-old twins' internalizing and externalizing problems. *Behavior Genetics, 33*, 409–420.
Van Hulle, C. A., Rodgers, J. L., D'onofrio, B. M., Waldman, I. D., & Lahey, B. B. (2007). Sex differences in the causes of self-reported adolescent delinquency. *Journal of Abnormal Psychology, 116*, 236–248.
Wade, T. D., & Kendler, K. S. (2000). The genetic epidemiology of parental discipline. *Psychological Medicine, 30*, 1303–1313.
Zheng, Y., Rijsdijk, F., Pingault, J. B., McMahon, R. J., & Unger, J. B. (2016). Developmental changes in genetic and environmental influences on Chinese child and adolescent anxiety and depression. *Psychological Medicine, 46*, 1829–1838.

Nicole Barbaro has a PhD in psychology with a specialization in evolution and human development. She is currently an Adjunct Professor of Psychology at Utah Valley University and Senior Communications Content Manager at WGU Labs. Nicole has published more than 40 scholarly articles on close relationships and human development. You can learn more at www.nicolebarbaro.com.

Eric J. Connolly is an Associate Professor in the Department of Criminal Justice and Criminology at Sam Houston State University. His research examines biological and social influences on antisocial behavior and victimization across the life course. His work has appeared in journals such as *Child Development*, *Child Maltreatment*, *Criminology, Journal of Quantitative Criminology*, and *Social Science & Medicine*.

Madi Sogge received her master's degree in criminology and criminal justice, with a focus on emergency management and biosecurity from the Saint Louis University College for Public Health and Social Justice. Her research interests center around the biological influences on human behavior and the intergenerational transfer of intimate partner violence. Madi has previously worked with survivors of domestic violence and is currently working as a registered nurse.

Todd K. Shackelford is Distinguished Professor and Chair of the Department of Psychology at Oakland University in Rochester, Michigan. Much of Shackelford's research addresses sexual conflict between men and women, with a focus on men's physical, emotional, and sexual violence against their intimate partners. Since 2006, Shackelford has served as editor of the journal *Evolutionary Psychology* and in 2014 founded the journal *Evolutionary Psychological Science* as Editor-in-Chief. Shackelford is an elected Fellow of the Association for Psychological Science and the American Psychological Association.

Brian B. Boutwell is an Associate Professor in The Department of Criminal Justice and Legal Studies in The School of Applied Sciences at The University of Mississippi. He also holds a secondary appointment as Associate Professor and Scientist Educator in The John D. Bower School of Population Health at the University of Mississippi Medical Center. His research focuses heavily on the use of twin and family designs to help improve causal inference abilities in observational data. In particular, this involves exploring possible causal pathways leading to various deleterious outcomes including disease, diminished health, psychopathology, and antisocial behaviors.

Authors and Affiliations

Nicole Barbaro[1] · Eric J. Connolly[2] · Madi Sogge[3] · Todd K. Shackelford[1] · Brian B. Boutwell[4,5]

✉ Nicole Barbaro
nmbarbar@oakland.edu

Eric J. Connolly
ejc025@shsu.edu

Madi Sogge
madi.sogge@slu.edu

Todd K. Shackelford
shackelf@oakland.edu

Brian B. Boutwell
bbboutwe@olemiss.edu

[1] Department of Psychology, Oakland University, 654 Pioneer Drive, Rochester, MI 48309, USA

[2] Department of Criminal Justice and Criminology, Sam Houston State University, Huntsville, TX 77340, USA

[3] Criminology and Criminal Justice, Saint Louis University, St Louis, MO 63103, USA

[4] Department of Criminal Justice & Legal Studies, The University of Mississippi, 84 Dormitory Row West, University, MS 38677, USA

[5] John D. Bower School of Population Health, University of Mississippi Medical Center, 84 Dormitory Row West, University, MS 38677, USA