

Is the Link Between Intelligence and Age at First Intercourse Causal or Correlational?
 A Cross-Generational Sibling Comparison Design Using the NLSY

Anecdotal evidence from the popular media, such as MTV’s reality television franchise, *16 and Pregnant*, suggests that teenage sexual involvement is on the rise. Academic research supports such evidence; age at first intercourse (AFI) is indeed declining and has for some time (Bozon, 2003; Finer, 2007; Kann et al., 2014). Early AFI is associated with downstream consequences, including lower education attainment (Harden, 2012; Spriggs & Halpern, 2008; Wellings et al., 2001), failure to meet education and career goals (Halpern, Joyner, Udry, & Suchindran, 2000), increased risk of teenage pregnancy (Leitenberg & Saltzman, 2000; Wellings et al., 2001), and increased rates of sexually transmitted infections (Kaestle, Halpern, Miller, & Ford, 2005). Moreover, beyond the obvious benefit of avoiding those negative outcomes, delaying AFI is associated with greater relationship satisfaction, perception of increased attractiveness, and higher household income (Harden, 2012). Because many of the negative consequences above are severe and long-reaching, it is important to identify the causal mechanisms associated with early AFI. One potential causal influence on AFI that has received attention is intelligence.

Higher levels of intelligence are positively associated with delaying first intercourse (Halpern et al., 2000; Mott, 1983; Paul, Fitzjohn, Herbison, & Dickson, 2000; Woodward, Fergusson, & Horwood, 2001), as well as less-intimate sexual involvement (Halpern et al., 2000). Specifically, it has been suggested that intelligent individuals delay intercourse to “safeguard” their futures (Kirby, 2002b; Manlove, 1998; Raffaelli & Crockett, 2003). They perceive the risks associated with early intercourse, (*e.g.*, pregnancy, STIs) to have life- and career-shattering outcomes (Halpern et al., 2000; Harden & Mendle, 2011). Although the link between intelligence and AFI has face validity, and has been confidently asserted (or often implied) as a causal link, a fundamental confound exists in most past research that limits our ability to infer causality.

Virtually all of the AFI-intelligence literature has used between family designs. In

all such analyses, genetic and environmental influences, such as education and maternal intelligence are confounded (D’Onofrio, Lahey, Turkheimer, & Lichtenstein, 2013; Harden, 2014; Lahey & D’Onofrio, 2010; Rodgers, Cleveland, van den Oord, & Rowe, 2000). By ignoring such confounds, the source of variance is ambiguous, and researchers that attribute the source to specific between- or within-family sources risk misattributions of causality (Rowe & Rodgers, 1997; Rutter, 2007). Indeed, both intelligence and AFI are (Harden & Mendle, 2011; Harden, 2014; Plomin & Spinath, 2004; Rodgers, Rowe, & Buster, 1999; Rodgers, Rowe, & May, 1994). Thus, we need to critically evaluate whether intelligence is a cause of AFI or merely a theoretically attractive confound. To resolve some of these methodological challenges, we use design innovations that emerge from the excellent cross-generational and longitudinal structure of the National Longitudinal Survey of Youth (NLSY; we use both the original NLSY79 survey and the NLSY-Children survey, described below).

Cause or Confound?

There are numerous theories that address the motivations for adolescents’ initiation of first intercourse (see Rodgers, 1996 or Buhi & Goodson, 2007 for reviews), and even more specific precursors to first intercourse (Buhi & Goodson, 2007; D’Onofrio & Lahey, 2010; Kirby, 2002a; B. C. Miller et al., 1997; Santelli & Beilenson, 1992). Many of these theories emphasize biology/genetics, where adolescent pubertal development (and associated hormone changes) drives the onset of sexual behavior (W. B. Miller et al., 1999; Udry, 1979, 1994). Other theoretical frameworks use social/environmental processes to explain developing sexual involvement in adolescence, such as Social Learning (DiBlasio & Benda, 1990; Hogben & Byrne, 1998), where social norms affect the likelihood of early sexual behavior; or Social Control theory (Hirschi, 2002), where societal and cultural influences reduce the likelihood that individuals will act on their natural tendency toward sexual involvement. Under these environmental theories the underlying biology is either ignored or actively resisted (*e.g.*, Social Control theory), while under many of the biological/genetic theories, the environmental

components are often ignored.

However, numerous articles have also advocated integrative models (See Harden, Mendle, Hill, Turkheimer, & Emery, 2008; Harden, 2014; Udry, 1995). The integrative Biopsychosocial Model acknowledges both genetic and environmental contributions to human behavior (Engel, 1977; Petersen, 1987; Rodgers et al., 1999). Indeed, biology, psychology, and society jointly influence adolescents' decisions to engage in sexual intercourse (Meschke, Zweig, Barber, & Eccles, 2000; Zimmer-Gembeck & Helfand, 2008).

Intelligence as Cause

The short-term risks of early AFI are primarily negative, whereas the rewards for delay are primarily positive. These consequences are suggested to extend into adulthood – early AFI has been related to adult delinquency (Harden et al., 2008), anti-social behavior, and substance abuse (Boislard & Poulin, 2011), whereas those with delayed AFI have higher household incomes in adulthood (Harden, 2012). It is intuitively appealing to believe that intelligent individuals are more likely to observe this potential risk-reward trade off, and through volition act upon such observations by delaying first intercourse. Accordingly, intelligent individuals perceive the consequences of early AFI to negatively influence their careers (Halpern et al., 2000; Harden & Mendle, 2011).

Indeed, most of the literature has contributed to this belief, that intelligence is causally connected to AFI. Those with higher educational goals delay their first intercourse (Boislard & Poulin, 2011; Schvaneveldt, Miller, Berry, & Lee, 2001), whereas those who engaged in early sexual intercourse reduced their educational goals compared to earlier higher goals (Schvaneveldt et al., 2001). Beyond academic goals, those with a greater affinity for risk and those who perceive benefits from teen-pregnancy are more likely to engage in risky sexual activities (Raffaelli & Crockett, 2003). A greater understanding of the risks associated with sexual intercourse, such as HIV transmission, is also associated with delayed AFI (C. Mathews et al., 2009).

Smarter adolescents are more likely to report delayed intercourse (Halpern et al.,

2000; Mott, 1983; Paul et al., 2000; Woodward et al., 2001). Beyond intercourse, smarter individuals appear to postpone all sexual/romantic activity (Halpern et al., 2000). Such blanket delays may be a proactive attempt to avoid first intercourse precursors. Thus, many researchers have concluded that “[h]igher intelligence operates as a protective factor against early sexual activity during adolescence, and lower intelligence, to a point, is a risk factor.” (Halpern et al., 2000, p. 213).

However, Halpern et al. (2000) and many of the other studies we have referenced above(e.g., C. Mathews et al., 2009; B. C. Miller et al., 1997; Paul et al., 2000) have used between family, typically cross-sectional, designs. Such designs cannot logically distinguish between processes that act to create differences between families and processes that create differences among family members (Lahey & D’Onofrio, 2010; Rodgers et al., 2000). Thus the previous studies do not provide conclusive evidence that intelligence is the causal influence behind the AFI-intelligence relationship. Logically, other alternatives are that AFI has a causal link to intelligence (which is unlikely, for the obvious theoretical reasons, including that a child’s intelligence precedes AFI in time) or that other confounds cause these two outcomes to correlate, but not causally. There are dozens, perhaps hundreds, of such confounds that can logically contend to explain the link between child intelligence and AFI.

Intelligence as a Confound

An equally valid set of explanations exist in which intelligence is not the causal factor behind the AFI-intelligence relationship, but rather one of dozens of potentially explanatory processes (though obviously a theoretically explanatory process). Instead, various confounds including family level selection effects, or third variables at the individual level could be causing the relationship. Indeed many such findings that link intelligence with various outcomes are quite possibly the product of misattributing between family confounds to individual-level causes.

The relationship between birth order and intelligence is a classic example of this misattribution (See Damian & Roberts, 2015; Rodgers et al., 2000; Rodgers, 2014).

Between-family studies that rely on cross-sectional data have consistently found that first born children have higher IQs than later born children (Belmont & Marolla, 1973; Zajonc, 1976). Yet within-family studies have often found zero relationship (Berbaum & Moreland, 1980; Retherford & Sewell, 1991; Rodgers et al., 2000, ; however, see Barclay (2015); Bjerkedal, Kristensen, Skjeret, & Brevik (2007) which have found small, but significant within-family effects in very large national studies). Moreover, when designs that can distinguish within- and between-family variance have been conducted, the methodological source of the IQ-birth order effects have emerged from the between-family variance (Black, Devereux, & Salvanes, 2011; Rodgers, 1984; Rodgers et al., 2000; Wichman, Rodgers, & Maccallum, 2006, 2007). Potential causes of this confound include maternal age at first birth, parental IQ, parental education, and SES (Page & Grandon, 1979; Rodgers, 2001; Rodgers et al., 2008, also see Anastasi (1956) for an insightful overview, written prior to the IQ-birth order debate.).

There is support in the literature suggesting that between-family influences such as SES and maternal intelligence could drive the relationship. Socioeconomic status is associated with the onset of first intercourse (Lammers, Ireland, Resnick, & Blum, 2000), explains much of the negative consequences of teenage pregnancy

Prior Within Family Analyses

Two past studies have explicitly separated between- and within-family influences on the AFI-intelligence relationship (Harden & Mendle, 2011; Nedelec, Schwartz, Connolly, & Beaver, 2012). Harden and Mendle used 536 same-sex twin pairs from the Add Health Study to “test[] whether relations between intelligence, academic achievement and age at first sex were due to unmeasured genetic and environmental differences between families.” Twins who differed in their intelligence or their academic achievement did not differ in their age at first intercourse. They concluded that “the association between intelligence and age at first sex could be attributed entirely to unmeasured environmental differences between families.” Nedelec, Schwartz, Connolly, and Beaver (2012) conducted an exploratory analysis of MZ twin pairs from the same

sample used by Harden and Mendle, using intelligence difference scores to predict various social outcomes. They found consistent null results, though their samples were small and their statistical analyses were substantially underpowered.

Current Study, Summary

To summarize, the current study examines the relationship between intelligence and age at first intercourse, using siblings and their children from a multi-generational nationally representative sample, the NLSY. This examination extends the intelligence literature in several key ways. First, we tested whether the relationship between intelligence and AFI existed in either or both between- and within-family analyses. Second, we evaluated the alternative explanation that maternal intelligence influences child AFI, using the cross-generational structure of the NLSY. Third, we replicated our findings using two different age periods.

We made the following predictions, based primarily upon Harden & Mendle (2011):

Between Families,

1. Does Gen2 intelligence predict Gen2 AFI?: We expect intelligence to be associated with age of first intercourse because there is a sizable body of literature reporting that result (Kirby, 2002b; Manlove, 1998; Raffaelli & Crockett, 2003).

2. Does Gen1 intelligence predict Gen2 AFI?: We also expect maternal intelligence to be associated with age of first intercourse because the heritability of intelligence is quite high (Bouchard, Jr., 2004; Devlin, Daniels, & Roeder, 1997). If intelligence does causally influence AFI we would expect that the cross-generational association between AFI and intelligence would be considerably weaker, but existent. However, if the intelligence-AFI relationship is the product of between family confounds, then we would expect that the cross-generational association between AFI and intelligence would be stronger than the within generation association because maternal intelligence would be more closely linked with household SES and various parental causes. Comparably sized effects would also be consistent with a between

family confound. Given that Harden & Mendle (2011) found no within family effect for intelligence, we expect that maternal intelligence will have a comparable or larger effect on between family AFI than child intelligence.

Within Families,

3. Does Gen2 intelligence predict Gen2 AFI?: No, we do not expect to find all within family differences in intelligence and AFI, given that Harden & Mendle (2011) did not report an effect.

4. Does Gen1 intelligence predict Gen2 AFI?: Unknown: it is possible that maternal intelligence will have an effect, as such a link would explain the between family effects as well as many of the alternative household-level influences.

5. Is the relationship consistent across methods?: Doubtful, we do not expect the results to be consistent across methods because both Harden & Mendle (2011) and Meredith (2013) found no within-family effect, while the traditional findings from between family studies find an effect (Kirby, 2002b; Manlove, 1998; Raffaelli & Crockett, 2003).

Method

Research Design

We adapted Kenny and colleagues (2001; 2006) reciprocal standard dyad model to facilitate sibling comparisons. Sibling-based quasi-experimental models are particularly effective for incorporating genetic and environmental design elements (Lahey & D’Onofrio, 2010; Rutter, 2007). Our model uses differences between both pairs of mothers and pairs of adolescents, which provides the measures that explicitly account for within-family variance. Further, within-family differences create a powerful control for virtually all background heterogeneity (variance) associated with both genetic and environmental differences (Lahey & D’Onofrio, 2010). We compare individuals from within the family in the context of the following models. First, we predict the difference in AFI, $Y_{i\Delta}$, for a given pair of NLSY-Children, indexed as i , in the following model:

$$Y_{i\Delta} = \beta_0 + \beta_1 \bar{Y}_i + \beta_2 \bar{X}_i + \beta_3 X_{i\Delta} \quad (1)$$

where,

$$Y_{i\Delta} = Y_{i1} - Y_{i2}; X_{i\Delta} = X_{i1} - X_{i2}, \text{ and} \quad (2)$$

$$Y_{i1} = \max(Y_{ij}); Y_{i2} = \min(Y_{ij}) \quad (3)$$

In this model, the relative difference in kin outcomes (Y_{Δ}) is predicted from the mean level of AFI (\bar{Y}), the mean level of intelligence, the predictor (\bar{X}), and the between-kin intelligence difference ($X_{i\Delta}$). The mean levels support causal inference through at least partial control for genes and shared environment in previous generations. Within this model, there is explicit separation of within-family variance (within Y_{Δ} and X_{Δ}), and between-family variance (within the \bar{Y} and \bar{X}).

Thus, this model allows us to explicitly untangle between- and within-family influences. If there is a true causal link between intelligence and AFI, then we expect kin differences in intelligence to be significantly associated with kin differences in AFI. If the effect is spurious – only the function of between-family confounds – then we would expect to find no significant relationship between the differences in the outcome with the differences in the predictor.

Sample

The National Longitudinal Survey of Youth 1979 dataset(NLSY79) is based on a nationally representative household probability sample, jointly sponsored by the US Bureau of Labor and US Department of Defense. On December 31, 1978, 12,686 adolescents were sampled within a household probability sample from 8,770 households on a battery of measures. The initial sample consisted of three subsamples:

- a cross-sectional household probability sample of 6,111 non-institutionalized adolescents residing in the United States on December 31st of 1978;
- a separate over-sampled civilian subsample of 5,295 racial minorities and disadvantaged whites;
- a representative sample of 1,280 youth serving in the US Military on September 30th, 1978.

In the two civilian samples, subjects' birthdates ranged from January 1, 1957 to December 31, 1964, and were between the ages of 14 and 21 on December 31, 1978; whereas military subject's birthdates ranged from January 1, 1957 to December 31, 1961, and were between 17 and 21 years old. Participants were surveyed annually until 1994, and then surveyed biennially to the present. Two waves of planned attrition occurred. After the 1984 interview, all but 201 randomly selected members of the military sample were dropped. After the 1990 interview, all 1,643 disadvantaged whites from the oversample were dropped. More information about the sampling process and the data can be found on the Bureau of Labor Statistics (BLS) website: <http://www.bls.gov/nls/nlsy79.htm>.

In 1986, the biological children of the female NLSY79 participants were surveyed, resulting in the NLSY79 Children and Young Adults (NLSY-Children) survey. These 11,512 participants are also surveyed on a biannual basis. Accordingly, participants in the NLSY79 will typically be referred to as the Generation 1 (Gen1) sample, whereas the NLSY-Children will be referred to as the Generation 2 (Gen2) sample.

Tetrads

To conduct our study using the necessary within-family information, we require sister pairs in Generation 1 who both had children. The children of these sisters are cousin-pairs. In the original NLSY79 and NLSY-Children surveys, identification of level of sibling relatedness in the NLSY was primarily inferential. NLSY79 twins, full siblings, half siblings, and adoptive siblings were distinguishable indirectly from respondent and maternal information about the biological father. NLSY-Children respondents within a given family were all full- or half-siblings, because they were (by design) the biological children of the NLSY79 females. In 2006, the NLSY surveys both included explicit indicators of the level of sibling relatedness. Our research team has recently completed a final multi-year project to reliability and validly identify the kinship pairs within these two datasets (Rodgers et al., 2015). These kinship pairs are used in the current study.

Specifically, Mother-Child-Aunt-Nibling (MCAN) tetrads were created using the

NLSY Kinship Links (Rodgers et al., 2015) and supporting R package (Beasley et al., 2015). The oldest two female kin (Mother, Aunt) were selected from each NLSY79 household (note that additional female Generation 1 sister pairs – a relatively small number – were excluded). Three tetrad designs were employed, in which the genders of Generation 2 were the defining feature:

- Mother-Daughter-Aunt-Niece (MDAN) tetrads included the oldest Generation 2 female child from each of the two Generation 1 sisters,
- Mother-Son-Aunt-Nephew (MSAN) tetrads included the oldest Generation 2 male child from each of the Generation 1 sisters, and
- The first two types of tetrads were combined together into Mother-Child-Aunt-Nibling (MCAN) included the first born child from each of the Generation 1 sisters. (Note: “Nibling” refers to an niece or nephew with unspecified gender; compare to “Sibling.”)

Measures

Generation 1 AFI. NLSY-79 subjects indicated their AFI a maximum of three times, in response to questions in 1983, 1984, and 1985. The 1984 and 1985 questions were included to assess those with non-response in 1983, but in fact many female respondents were surveyed multiple times. Further, female were asked for additional related information (Year of First Intercourse, Month of First Intercourse) in 1984 and 1985.

We used the repeated questions to estimate the test-retest reliability of self-reported AFI and AFI difference scores. In Table 1, the lower triangle reports the correlations of self-reported AFI across 1983-1985; the diagonal indicates the number of respondents reporting AFI for that year, and upper triangle indicate the number of respondents that reported AFI for both respect years. The test-retest correlations are high ($r > .75$) across all viable pairings, suggesting that our subjects are consistently reporting their AFIs.

Table 1

Correlation of Gen1 Self-Reported AFI Across 1983-1985

	1983	1984	1985
1983	8432	3765	88
1984	0.86	4516	0
1985	0.76	NA	424

Generation 2 AFI. Over the life-time of the NLSY-Children survey, participants were asked approximately the same questions to assess AFI that their mothers were asked. However, Generation 2 respondents were only asked for AFI information once they had reached age 15 or later. The exact nature of the question varied by administration. Between 1988 and 2000, subjects were asked for age, year, and month of first intercourse. After 2000, subjects were only asked their age.

We calculated AFI using a multi-step process for three reasons: (1) to account for the diversity of AFI questions across survey administrations, (2) to incorporate multiple reports by the same participant, and (3) to account for the imprecision of AFI reporting(*e.g.*, a subject who reports AFI at 16 could be any age between 16 years and 1 day old through 16 years and 364 days old).

For each survey, we transformed year of first intercourse into AFI. If subjects reported both age and year within the same survey and ages were different, we averaged the AFI scores. Across surveys, we identified the earliest AFI and the latest AFI for each subject. We designated these two AFIs as the Minimum AFI and the Maximum¹ AFI respectively, thus identifying the full range of possible AFIs for each participant. We used this AFI range to calculate the expected value of AFI. Using this method, the average Generation 2 AFI was 16.01 years (sd = 2.30; n = 6288).²

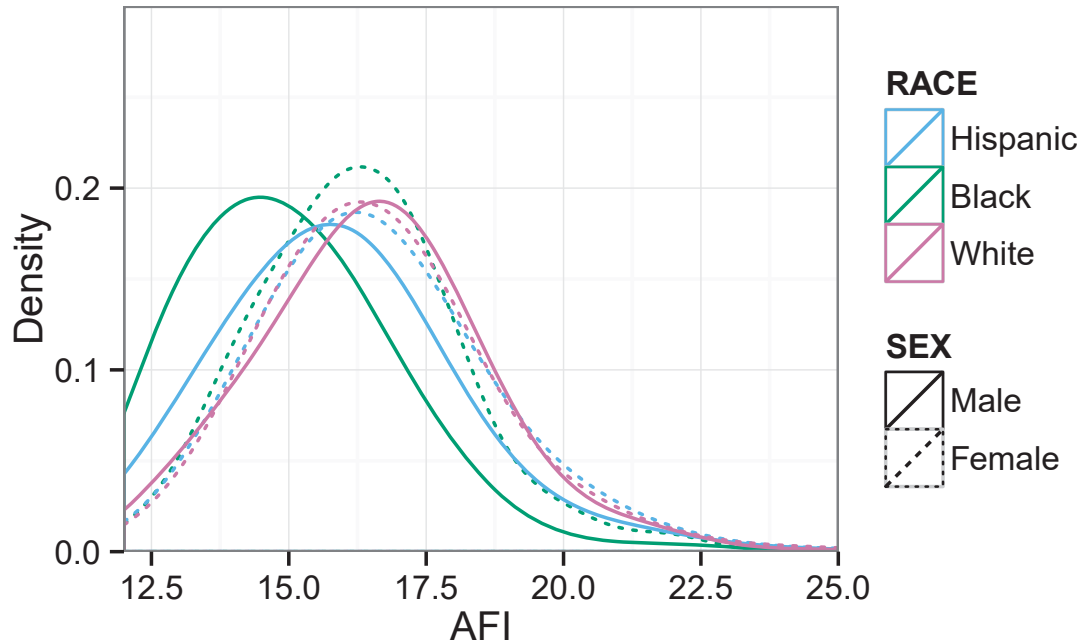
¹We added 1 year to the Maximum AFI to address the imprecision of self-reported age. The expected value of AFI of any subject does not equal the reported AFI. For example, a subject who reports AFI at 16 could be anywhere from 16 years and 1 day old to 16 years and 364 days old.

²Taking the average of all AFIs (without addressing expected value), results in 15.49 (sd = 2.30; n = 6288). Adding in expected value of .5 changes this value to 15.99.

My motivation was to incorporate the maximum amount of information into our calculated AFI, given that age reporting is imprecise (*i.e.* two people can have AFIs of 16, but one can have been 16 years and one day old and the other could have been 16 years and 364 days old. Thus the expected value for 16 is in fact 16.5). In essence, what we did was find the midpoint of possible AFIs for each subject. The minimum AFI was the earliest possible reported AFI across all survey waves. And the maximum was the latest reported AFI plus 1

Mason - I find this paragraph really confusing, both the motivation and the procedure - I helped you do this, and I still don't know what most of this means. Did we create a SAS variable for age when we could? And otherwise created the most precise variable we could with what we had? More discussion about this before we're ready for this paragraph to go public.

After transforming all AFI scores, we recoded impossible AFIs as missing. A score was impossible if the reported AFI that exceeded participant's age at time of survey ($\overline{\text{AFI}} = 15.99$, $\text{sd} = 2.30$, $n = 6235$). Next, we excluded all AFIs below age 12 ($\overline{\text{AFI}} = 16.14$, $\text{sd} = 2.10$, $n = 6087$). Finally we excluded subjects who reported AFI prior to menarche ($\overline{\text{AFI}} = 16.16$, $\text{sd} = 2.09$, $n = 6047$). We excluded those below age 12 because those responses likely are the result of misunderstanding or non-consensual sexual activity, while we excluded those with premenstrual AFI because of we were only interested in post-pubescent sexual activity. AFI varied by gender and race. Most notably, women reported AFIs that were 6 months later than men, and black men reported the lowest AFI (15 yrs) of any race-gender categories. For a complete portrayal of summary statistics, see Table 2 and see Figure 1.

Figure 1. Smoothed Density Plots of AFI by Sex and Race

Generation 1 Intelligence. The Armed Services Vocational Aptitude Battery (ASVAB; Form 8A; Palmer, Hartke, Ree, Welsh, & Valentine, Jr., 1988) was administered to Gen1 participants in 1980. The Armed Forces Qualification Test (AFQT) is contained within the ASVAB, and used as a measure of general trainability (Maier & Sims, 1986). It is a composite of four subscales: Arithmetic Reasoning (AR; 30 items), Math Knowledge (MK; 25 items), Paragraph Comprehension (PC; 15 items), and Word Knowledge (WK; 35 items). Other administrations of the pencil and paper ASVAB reveal that all the AFQT subscales have high coefficient α internal consistency ($\alpha_{AR} = .91$; $\alpha_{WK} = .92$; $\alpha_{PC} = .81$; $\alpha_{MK} = .87$; Kass, Mitchell, Grafton, & Wing, 1982). Reported reliability of the overall AFQT (version 8A) ranges from .87 to .93 (Palmer et al., 1988).

Methods of calculating the AFQT have varied throughout the ASVAB's administrative lifetime (Mayberry & Hiatt, 1992). For pencil and paper administrations, standard scores were created for each of the subscale scores ($\bar{x} = 50$, $sd = 10$), and then combined into a standard score. Then, the AFQT standard score is

derived from the following formula:

$$\text{AFQT} = \text{AR} + \text{MK} + 2\text{VE}, \quad (4)$$

$$\text{where, } \text{VE} = \text{PC} + \text{WK}. \quad (5)$$

Many researchers have used the AFQT standard score as a proxy for general intelligence (g) (Herrnstein & Murray, 1994; Der, Batty, & Deary, 2009). Indeed, the military has found that the AFQT correlated 0.8 with the Wechsler Adult Intelligence Scale (WAIS; McGrevy, Knouse, & Thompson, 1974). Moreover, the AFQT consistently predicts outcomes traditionally associated with intelligence (Welsh, Kucinkas, & Curran, 1990), including grades (Wilbourn, Valentine, Jr., & Ree, 1984; J. J. Mathews, 1977).

Generation 2 Intelligence. NLSY-Children respondents, as a part of the survey, complete the following test batteries:

- Peabody Individual Achievement Test (PIAT; Dunn & Markwardt, 1970):
 - Math Subtest (84 items),
 - Reading Recognition Subtest (84 items),
 - Reading Comprehension Subtest (84 items),
- The Peabody Picture Vocabulary Test-Revised (PPVT-R; Form L; Dunn & Dunn, 1981; 175 items), and
- Digit Span Subscale of the Wechsler Intelligence Scales for Children-Revised (Digit Span; Wechsler, 1974; 28 items).

The standard scores of the PPVT-R, PIATs, and Digit Span are considered valid and reliable assessments of cognitive ability (Mott & Baker, 1995). However, subjects were surveyed on a biennial basis. Thus we could not use cognitive tests at a fixed age. Instead, we aggregated scores across a 4 year window, and targeted the midpoint between ages 9 and 10. We targeted 9.5 because all cognitive tests were administered within the 8–11 age window, we wanted to maximize the number of subjects with viable ability scores, and we wanted to ensure temporal precedence by measuring intelligence

prior to the occurrence of AFI. In the case of missing subtests, we allowed age 11 standard scores to replace age 9 standard scores, and age 8 standard scores to replace age 10 standard scores. Our replacement strategy ensured that the average age of testing matched the average of our targeted ages. To obtain intellectual ability measures for each NLSY-children respondent, we fit a confirmatory factor analysis model (using Mplus; Muthén & Muthén, 2014) and their robust maximum likelihood estimator option. A single-factor model fit fairly well ($RMSEA = .101$; $CFI = .973$; $TLI = .946$), supporting the use of this model to construct a unidimensional scale score for each respondent. We used factor scores obtained from this model as our measures of NLSY-Children intelligence.

Replicability & Reliability. We repeated our aggregates of Gen2 intelligence, centered at ages 10.5 and 11.5, and replicated all of our analyses. These replications can be found in the Appendices A and B, respectively. The test-retest reliabilities of Gen2 intelligence across our three aggregations is reported in the lower triangle of Table 3. The diagonal indicates the number of respondents with intelligence aggregations for that year, and upper triangle reveal the number of respondents with viable scores for both respective ages. The test-retest correlations are very high ($r > .90$) across all pairings, suggesting that our method captures consistent measures of intelligence across ages.

Table 3

Correlation of Gen2 Aggregated Intelligence across Ages 9.5, 10.5, 11.5

	Age 9.5	Age 10.5	Age 11.5
Age 9.5	8254	7974	7669
Age 10.5	0.95*	8143	7838
Age 11.5	0.90*	0.96*	7970

Reliability of Difference Scores

Our design assumes that the difference scores of our measures are reliable. We've reported the test-retest reliability of Gen2 intelligence and Gen1 AFI in earlier sections. Here, we report the test-retest reliability of the pairwise differences of those measures.

AFI. Comparing sibling differences in AFI as reported in 1983 and 1984 ($n = 783$ pairs) we found a strong correlation ($r = .76$). The sample of sibling pairs with complete information in 1985 was too small ($n = 12$ pairs) to compare to the other two years. Regardless, sibling differences in self-reported AFI appear reliable. Because we could not estimate test-retest reliabilities for Generation 2, we calculated the reliability using Lord's equation (1963). Generation 2 Mean AFI difference scores were also reliable ($r = 0.734$) and comparable to Generation 1 sibling differences.

Intelligence. Cousin differences in intelligence as assessed at ages 9.5, 10.5, and 11.5 were correlated using three different linking methods (Mixed, Daughters, Sons). Reliabilities across linking methods was consistent and high (min $r = .86$; max $r = .95$). However, we could not estimate test-retest reliabilities for Generation 1; we calculated the reliability using Lord's equation (1963). The calculated reliability of Generation 1's differences in AFQT was 0.606, acceptable, but lower than the empirical correlation we derived for cousin differences.

Results

We examined the relationship between AFI and intelligence using two designs: a between-family design, and a within-family design (which also includes between-family variance in the differences between the family means). The results are organized into those two designs. The between-family analyses report the relationships between the within-family average AFI and various measures of ability. The within-family analyses replicate the between-family findings using within-family difference scores by testing whether differences in AFI can be explained by differences in various measures of ability.

Between-Family Analyses

First, we examined the between family results. We tested whether the family average of Gen2 AFI could be predicted by the family averages of Gen1 intellectual ability and of Gen2 intellectual ability. We evaluated the influences both independently and simultaneously. All intelligence scores have been standardized by generation ($\bar{g} = 0$,

sd = 1), prior to averaging by household. AFI scores have been standardized by gender ($\overline{\text{AFI}} = 0$, sd = 1), prior to averaging by household.

Gen1 Mean Intelligence → Gen2 Mean AFI. Gen1 sister averages (NLSY79 mothers) of standardized AFQT scores were used to predict Gen2 averages of gender standardized AFI. Table 4 displays the results by Gen2 linking. The Mixed model reports the averages of the first born child (both males and females) of each sister (n = 342), the Daughters model reports the averages of the first born girls (n = 264), and the Sons model reports the averages of the first born sons (n = 282). All three models reveal similar results. A one unit increase in the average standardized intelligence of the children's mothers predicted a statistically significant increase of $\approx .013$ standard deviations in average Gen2 AFI. The adjusted R^2 varied slightly by Gen2 linking (Mixed = .087, Daughters = .097, Sons = .103).

Gen2 Mean Intelligence → Gen2 Mean AFI. Gen2 cousin averages of standardized intellectual ability scores were used to predict Gen2 averages of gender-standardized AFI. Table 5 displays the results by Gen2 linking. The Mixed model reports the averages of the first-borns of each of the NLSY79 mothers (sisters) (n = 344), the Daughters model reports the averages of the first born girls (n = 267), and the Sons model reports the averages of the first born sons (n = 283). All three models reveal similar results. A one unit increase in the average standardized intelligence of the children predicted a statistically significant $\approx .075$ standard deviation increase in average Gen2 AFI. The adjusted R^2 varied slightly by Gen2 linking (Mixed = .014, Daughters = .016, Sons = .009).

double checked. These R^2 are correct.

Joint Mean Intelligence → Gen2 Mean AFI. Results from the Gen1 sister averages of standardized AFQT scores and Gen2 cousin averages of standardized intellectual ability scores predictions of Gen2 averages of gender standardized AFI are displayed in Table 6. Again, three models based on Gen2 linking are displayed: Mixed (n = 337), Daughters (n = 260), and the Sons (n = 278). All three models reveal similar results. Gen1 (maternal) intelligence was significantly associated with Gen2 AFI (p

$< .01$), while Gen2 (child) intelligence was not significantly associated with Gen2 AFI. A one unit increase in the average standardized intelligence of the children's mothers predicted $\approx .014$ standard deviation increase in average Gen2 AFI. The adjusted R^2 varied slightly by Gen2 linking (Mixed = .086, Daughters = .097, Sons = .100); each was practically identical to the Mean Gen1 Intelligence models.

Within-Family Analyses

We replicated the between-family analyses reported in the previous subsection, using within-family difference scores and means. Using the discordant sibling model, we predicted the differences in Generation 2 AFI as a function of differences in intelligence, controlling for means of the outcomes and predictors. We ran three series of models, where we examined the individual and then joint influence of Gen1 intelligence and Gen2 intelligence. Moreover, within each series we included three Generation 2 linking method variants, just as we did in the between family analyses: Mixed model reports the differences of the first borns of each sister, the Daughters model reports the differences of the first born girls, and the Sons model reports the differences of the first born sons.

Gen1 Intelligence Differences \rightarrow Gen2 AFI Differences. Generation 1 sister differences in standardized AFQT scores were used to predict Gen2 differences of gender standardized AFI, controlling for Generation 1 sister averages of standardized AFQT scores and Gen2 averages of gender standardized AFI. Table 7 displays the results by Generation 2 linking method. The Mixed model reports the averages and differences of the first borns of each sister ($n = 336$), the Daughters model reports the averages and differences of the first born girls ($n = 258$), and the Sons model reports the averages and differences of the first born sons ($n = 278$). All three models reveal similar results. Generation 2 averages of gender standardized AFI (between-family measures) were significant predictors of Gen2 differences in gender standardized AFI ($p < .01$), across all three linking methods. A one unit increase in the average gender standardized AFI predicted ≈ 0.34 standard deviation increase in average Gen2 AFI difference, controlling for all over variables in the model.

In the Sons model, the Generation 1 sister average of standardized AFQT scores was a significant predictor of differences in Gen2 AFI ($p < .01$). A one unit increase in the average standardized intelligence of the children's mothers predicted $\approx .0083$ decrease in the AFI difference between siblings. All other variables were not significant, including all kin difference variables (the within-family measures). . The adjusted R^2 varied slightly by Generation 2 linking method (Mixed = .066, Daughters = .072, Sons

= .106).

Gen2 Intelligence Differences → Gen2 AFI Differences. Gen2 cousin differences in standardized intellectual ability scores were used to predict Gen2 differences of gender standardized AFI, controlling for Gen2 cousin averages of standardized ability scores and gender standardized AFI (to account for between-family variance). Table ?? displays the results by Generation 2 linking method. The Mixed model reports the averages and differences of the first borns of each sister ($n = 291$), the Daughters model reports the averages and differences of the first born girls ($n = 223$), and the Sons model reports the averages and differences of the first born sons ($n = 238$). All three models reveal similar results. Gen2 averages of gender standardized AFI were significant predictors of Generation 2 differences in gender standardized AFI ($p < .01$), across all three linking methods. A one unit increase in the average gender standardized AFI predicted ≈ 0.38 standard deviation increase in average Gen2 AFI difference, controlling for all over variables in the model.

In the Sons model, the Generation 2 cousin average of standardized ability scores was a significant predictor of differences in Generation 2 AFI ($p < .05$). A one unit increase in the average standardized intelligence of the children predicted $\approx .107$ decrease in the AFI difference between siblings. All other variables were not significant, including all kin difference variables. The adjusted R^2 varied slightly by Generation 2 linking method (Mixed = .103, Daughters = .121, Sons = .132).

Joint Intelligence Differences → Gen2 AFI Differences. Generation 1 sister differences in standardized AFQT scores and Gen2 cousin differences in standardized intellectual ability scores were used to predict Generation 2 differences of gender standardized AFI, controlling for Generation 1 sister averages of standardized AFQT scores, Gen2 cousin averages of standardized ability scores, and Gen2 cousin averages of gender standardized AFI. Table 9 displays the results by Generation 2 linking method. The Mixed model reports the averages and differences of the first borns of each sister ($n = 285$), the Daughters model reports the averages and differences of the first born girls ($n = 217$), and the Sons model reports the averages and differences of

the first born sons ($n = 235$). All three models reveal similar results. Gen2 averages of gender standardized AFI were significant predictors of Generation 2 differences in gender standardized AFI ($p < .01$), across all three linking methods. A one unit increase in the average gender standardized AFI predicted ≈ 0.38 standard deviation increase in Generation 2 AFI difference, controlling for all over variables in the model.

All other variables were not significant, including all kin difference variables. The adjusted R^2 varied slightly by Generation 2 linking method (Mixed = .090, Daughters = .105, Sons = .131).

Discussion

This article presents the relationship between AFI and intelligence using two different designs: a between-family design, and a within-family design (that also includes between-family variance within it as well). The between-family design allowed us to replicate previous researchers who used a cross-sectional sample. The within-family design allowed us to separate within- and between-family variance, to determine the source of the explanatory processes. The logic of this separate allows us to get much closer to the evaluating intelligence differences within the family to address issues of causality. The results revealed a stark contrast between the two methods, and cast doubt on the validity of past causal assertions.

Between- vs. Within-Family Variance

Between-Family Results. Notably, the between-family analyses showed a relationship between intelligence and AFI. Thus, we were able to replicate the findings of various researchers (Halpern et al., 2000; Mott, 1983; Paul et al., 2000; Woodward et al., 2001), and confirm hypotheses 1 and 2. However the relationship between AFI and intelligence was substantially stronger between maternal intelligence and child AFI than between the child’s own intelligence and child AFI, which suggests that family-level variables rather than individual-level intelligence is the likely source of the relationship. If the child’s own intelligence had been the primary causal influence on AFI we would have expected a considerably weaker cross-generational association between AFI and intelligence. Instead we find that the within-generation association is the weaker effect, suggesting that the child’s own AFI is likely derivative of the child’s mother’s intelligence, which is the more likely causal influence.

Thus, the “new” and alternative interpretation of this finding would be that maternal intelligence is driving the effect, and that past between-family analyses finding a link between child’s intelligence and AFI are likely because child’s intelligence is indirectly measuring maternal intelligence. Smarter mothers might be more effective at encouraging their children to delay intercourse – perhaps by effectively conveying the

riskiness of sexual intercourse (Hutchinson, Jemmott, Jemmott, Braverman, & Fong, 2003; C. Mathews et al., 2009). Considering that intelligence is highly heritable (Bouchard, Jr., 2004) and thus highly correlated across generations, this alternative explanation would still be consistent with the traditional between family findings, which do not control for maternal intelligence (Halpern et al., 2000; Mott, 1983; Paul et al., 2000; Woodward et al., 2001). However, we note that the Harden & Mendle (2011) findings, using biometrically-informed data, implicated the shared environment in this causal process. Our results are entirely consistent with theirs, using a different dataset and a different methodological approach to identify important sources of variance.

Within-Family Results. In the within-family analyses, the effect vanishes for both maternal intelligence and child intelligence. The child of the smarter Generation 1 mother was not more likely to delay intercourse compared to the child of the less-smart Generation 1 mother. Moreover, in spite of our finding that the Generation 1 intelligence was a relatively stronger predictor of Generation 2 AFI, we did not find that differences in Generation 1 intelligence to be associated with differences in Generation 2 AFI.

Concluding Remarks. These results cast doubts on the alternative explanation for the between-family results we posed in the previous paragraph. For, if Generation 1 maternal intelligence was driving the effect, we would have expected to find a significant within-family link from maternal intelligence to child AFI, which we did not. Rather, we think maternal and child intelligence are indirect measures of many other household features, any one of which may be more proximal as the causal explanation - income, parental education, family interaction. Or, the whole package of these features may stand in for a general environmental factor, a “little e,” which indexes the quality of home environment - a composite of parental income, intelligence, education, family interaction.

We interpret these results in relation to two previous findings. First, Rodgers et al. (2008) used Danish twin data, and found that the link from education/cognitive ability to maternal age at first birth was entirely accounted for by between-family variance:

“variance in AFB emerges from [IQ and education] differences between families, not differences between sisters within the same family” (Rodgers et al., 2008, p. S202). We have exactly the same result in the current study. Second, Harden and Mendle’s (2011) results, obtained using the Add Health data, use intelligence as a predictor and AFI as an outcome, just as we did. Their biometrical finding of meaningful shared environmental is the equivalent biometrical result to our finding of only between-family variance. But our design allows us to cast further doubt on the direct and causal influence of maternal intelligence, and leaves the general family environment and other factors correlated with maternal intelligence as the likely causal influences on AFI.