Social Science Research 43 (2014) 184–199

Contents lists available at ScienceDirect

Social Science Research

journal homepage: www.elsevier.com/locate/ssresearch

What can we learn from twin studies? A comprehensive evaluation of the equal environments assumption

Jacob Felson

William Paterson University, 300 Pompton Road, Wayne, NJ 07470, United States

ARTICLE INFO

Article history:
Received 13 August 2012
Revised 30 September 2013
Accepted 11 October 2013
Available online 22 October 2013

Keywords:
Behavior genetics
Twin studies
Equal environments assumption

ABSTRACT

Twin studies are a major source of information about genetic effects on behavior, but they depend on a controversial assumption known as the equal environments assumption (EEA): that similarity in co-twins' environments is not predictive of similarity in co-twin outcomes. Although evidence has largely supported the EEA, critics have claimed that environmental similarity has not been measured well, and most studies of the EEA have focused on outcomes related to health and psychology. This article addresses these limitations through (1) a reanalysis of data from the most cited study of the EEA, Loehlin and Nichols (1976), using better measures, and through (2) an analysis of nationally representative twin data from MIDUS using more comprehensive controls on a wider variety of outcomes than previous studies. Results support a middle ground position; it is likely that the EEA is not strictly valid for most outcomes, but the resulting bias is likely modest.

1. Introduction

Over the last four decades, behavior geneticists have found evidence that genes influence nearly all human behavior (Turkheimer, 2000; Freese, 2008). For outcomes like personality and educational attainment, researchers have found that the explanatory power of genes exceeds that of parental socialization (Rowe, 1995; Harris, 1999; Nielsen, 2006) (but see Nielsen and Roos, 2011). Conclusions such as these strike at the core of the sociological perspective, which maintains that the causal power of cultural forces far exceeds the causal power of genes. Although an increasing number of sociologists have integrated behavior genetic perspectives into their work (Guo and Stearns, 2002; Nielsen, 2006; Guo et al., 2007; Freese, 2008; Adkins and Vaisey, 2009; Conley et al., 2013), the idea that genes have a major influence on social behavior has not yet entered into the mainstream sociology curriculum. It is probably safe to say that many, if not most, sociologists remain skeptical that the effects of genes on social behavior are strong enough to warrant a fundamental shift in outlook.

Whether this skepticism is warranted depends to a large extent on the validity of twin studies, which provide much of the evidence for the importance of genetic effects on social behavior.[1] In the classic twin study (CTS), data are collected from monozygotic (MZ) twins, who are virtually identical genetically, and from DZ twins, who share about 50% of their segregating genes on average.[2] Similarity on a given trait is estimated, typically via correlation, for the MZ twins and for the DZ twins. When the correlation between outcomes of co-twins is higher among MZs than among DZs, a genetic effect on the outcome is inferred.

A key point of contention here is that genetic effects derived from twin studies may be biased upward if MZ co-twins share more similar environments in a way that induces greater similarity between co-twins on the outcome in question (Horwitz et al., 2003; Joseph, 2010; Conley et al., 2013). Although it is well-known that MZs experience more similar

E-mail address: felsonj@wpunj.edu

[1] Evidence for genetic effects also comes from adoption studies and from molecular genetic studies, both of which I discuss later.

[2] Segregating genes are genes that vary within the human species.

0049-089X/$ - see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.ssresearch.2013.10.004

environments than do DZs (Holmes, 1930; Wilson, 1934; Mowrer, 1954; Smith, 1965), there is little evidence that environmental similarity contributes to similarity in outcomes (Kendler et al., 1993; Conley et al., 2013). For this reason, behavior geneticists have generally held that twin studies are unbiased by environmental similarity between co-twins. This assertion is known as the Equal Environments Assumption (EEA), and it is disputed by critics who argue that the measurement of environmental similarity suffers from low validity and low reliability (Pam et al., 1996; Horwitz et al., 2003; Richardson and Norgate, 2005; Joseph, 2010).

This paper evaluates the EEA in a comprehensive manner and improves on previous research in at least three ways. First, I address concerns about reliability of measurement by estimating the reliability for each measure of environmental similarity that I use and considering how random error in measurement affects the results. Second, I address concerns about validity by measuring environmental similarity in a more comprehensive way than previous researchers have done. Third, unlike previous analyses which generally focused on a small number of outcomes within a particular subfield, in my main analysis, I examine a wide range of outcomes. By examining a range of disparate outcomes within a single dataset, I am able to discern whether environmental confounding of genetic effects is greater for some types of outcomes than it is for others.

The article is organized as follows. First, I explain why the results of twin studies are still worth debating today in an era of molecular genetics. Next, I review the rationale by which researchers make inferences about the effects of genes based on comparisons of monozygotic (MZ) and dizygotic (DZ) twins. Then I review previous research that has tested the equal environments assumption for specific traits. Included in this review is a reanalysis of data used in the most cited evaluation of the EEA, Loehlin and Nichols (1976). Finally, I test the EEA with respect to a variety of outcomes using a nationally representative sample of twins.

2. Why focus on twin studies in an age of molecular genetics?

In the classic twin study, phenotypic variation is parsed into environmental and genetic components. Some argue that this approach is no longer sensible in light of recent discoveries indicating that genetic effects are much more complex and contingent than previously believed (Charney, 2012). Though it was once thought that particular genetic variants (i.e. SNPs) might individually have a substantial impact on variation in complex phenotypes, a search has revealed very few strong, replicable effects (Manolio et al., 2009; Chabris et al., 2012). In addition, research has shown that heritability is not only about DNA; aspects of the biochemical system that regulate genetic expression, known as the epigenome, are also heritable (Charney, 2012). In light of these findings, why is it useful to validate an assumption underlying twin studies?

One reason it is important to evaluate the EEA is to help understand why estimates of genetic effects from twin studies are large when the effects of any particular SNP are small (Manolio et al., 2009). Part of the answer seems to be that the effects of genetic variants, while individually small, cumulate into larger effects. Using data on hundreds of thousands of SNPs identified in DNA samples from several thousand people, researchers have found more direct evidence for substantial genetic influence. They have found that the proportion of shared SNPs among a group of people correlates with phenotypic variation in that group on a variety of traits such as general intelligence (Davies et al., 2011; Chabris et al., 2012), policy preferences, education (Benjamin et al., 2012), neuroticism and extraversion (Vinkhuyzen et al., 2012). To some extent, this evidence supports the overall conclusion of twin studies that genes exert non-negligible effects on complex behaviors. On the other hand, estimates of cumulative genetic influence using molecular-level data have tended to be substantially lower than the corresponding estimates from twin studies. For example, a recent estimate of the proportion of variance in educational attainment explained by genes from twin studies was 0.35, whereas the corresponding estimate from molecular-level data was 0.16 (Benjamin et al., 2012). What accounts for the discrepancy? Since the SNP data that was collected did not capture all genetic variation, it is possible that studies using data at the molecular-level underestimate the effects of genes. Alternatively, twin studies may overestimate genetic effects due to violations of the EEA. A comprehensive examination of the EEA can help resolve this question.

3. Estimating heritability from twin studies

Before reviewing previous research on the EEA, I explain how twin studies can provide estimates of genes' explanatory power for a given trait provided that the equal environments assumption is valid. I consider a simple model in which there is no effect of assortative mating and no gene-environment interaction. The analysis begins by estimating correlations between co-twins on the trait of interest separately for MZ and DZ pairs. Because MZ twins share 100% of their genes and DZ twins share on average 50% of their segregating genes, co-twin correlations (r_{MZ} and r_{DZ}) can be decomposed into a heritability component (h^2) and a shared environmental component (c^2), as shown in Eqs. (1) and (2). If the equal environments assumption (EEA) is true, then $c^2_{MZ} = c^2_{DZ}$ and h^2 can be estimated easily by subtracting Eq. (2) from Eq. (1).[3] If on the other hand the EEA is false, then $c^2_{MZ} \neq c^2_{DZ}$ and estimates of h^2 will tend to be biased upward in twin studies.

[3] Estimation of heritability and shared environmental effects is usually accomplished with a structural equation model, but the logic of those models is analogous to the logic of the simpler models shown here.

$$r_{MZ} = h^2 + c^2_{MZ} \quad (1)$$

$$r_{DZ} = .5h^2 + c^2_{DZ} \quad (2)$$

Following the logic of Eqs. (1) and (2), researchers have estimated substantial genetic effects using twin data for dozens of attitudes and behaviors of interest to social scientists (for a review, see Freese (2008)). The validity of much of this research rests on the equal environments assumption.

4. Reassessing the most cited evaluation of the EEA

The most cited evaluation of the EEA was conducted by Loehlin and Nichols (1976) (L&N), who analyzed data on twins who had taken the National Merit Scholar Qualifying Test (NMSQT) as high school juniors in 1962. Data were collected from the twins about a wide variety of characteristics, including intellectual aptitude and personality. Also included were parents' reports of how similarly they treated their twin children, as well as twins' own reports about how similarly they felt they were treated. Loehlin and Nichols (1976) examined correlations between differences in treatment of the co-twins as reported by parents' and differences in outcomes of the co-twins. The set of correlations that they found was about what one would expect from random fluctuation around a low average value (about 0.035). L&N concluded it was unlikely that environment explained any more than a tiny fraction of the difference in outcomes between twins, and thus that the EEA was essentially valid.

Joseph (2010) challenged L&N conclusion, arguing that parents' reports about similarity in treatment might not be valid since parents may be loathe to acknowledge differences in how they treat their children. If Joseph is right, then perhaps the correlations between outcome similarity and *twins' own* reports of treatment similarity would be larger and more often significant than the correlations with parents' reports as originally reported by L&N.[4]

In an ancillary analysis, I examined this possibility using L&N's original data. Outcomes in this analysis included vocational interests, personality characteristics, the quality of interpersonal interaction, and test scores in English, math, social studies, natural science and vocabulary. Measures of environmental similarity used by L&N included reports from parents about the extent to which they treated their twin children similarly as well as whether their twins: played together between the ages of 6 and 12, tended to spend their time together between ages 12 and 18, had the same teacher in school, and slept in the same room. Measures of environmental similarity reported by twins (measured but not analyzed by L&N) included questions about the extent to which the twins wore similar clothes, shared friends, spent time together, and shared possessions with their co-twin.

I first replicated L&N's analysis by correlating the six parent-reported measures of environmental differences, as well as a composite of these measures, with absolute differences between co-twins on each of 35 outcomes separately for MZ twins and DZ twins (for a total of $7 \times 41 \times 2 = 574$ correlations). Correlations ranged between −0.16 and 0.24, with an average of about 0.037. Results mirrored what L&N found and were about what we would expect from chance fluctuation around a very small true correlation.

When I re-estimated these correlations using twins' own reports of environmental differences, I found correlations ranging between −0.09 and 0.27, with an average of about 0.06. So the effects of differences in treatment on differences in outcomes appear somewhat larger when treatment dissimilarity is measured by twins' own reports. Focusing only on the composite of twin reports, which should have higher content validity and higher reliability than individual measures, I found an average correlation of about 0.09.[5]

Confounding was not patterned across outcomes. For example, there was little consistency across zygosity in the magnitudes of correlations between the composite measures of environmental similarity and absolute differences in outcomes. A plausible conclusion is that nearly all outcomes examined by L&N are confounded to a small degree and that this confounding is modest enough to go undetected in many cases, even when environmental similarity is measured in a more comprehensive, reliable way. This conclusion is somewhat different from that reached by L&N, who acknowledged the existence of weak effects but concluded they were inconsequential. When environmental similarity is measured in a more reliable, accurate way, its effects appear large enough to be of some consequence for heritability estimates.

5. Overview of the literature testing the EEA

I identified 24 additional studies that focused on evaluating the EEA for some set of outcomes. These studies and the samples on which they were based (as well as the information about L&N's study) are listed in Table 1. Most of the samples used were from population registries or otherwise representative of national populations.

Table 2 lists measures of environmental similarity between co-twins used in studies of the EEA. The most commonly used measures of environmental similarity included perception of zygosity (by twins and/or parents), frequency of contact in youth and similarity of appearance. Less common measures included frequency of contact in adulthood and similarity of treatment by parents or by others during youth. Only a few studies examined whether twins' emotional closeness

[4] We would also expect stronger effects of averaged twin reports simply because measures based on averages of two reports will tend to be more reliable than measures based on a single report.

[5] An anonymous reviewer suggested that I examine composites.

Table 1
Studies testing the equal environments assumption with information about sample.

1	Scarr (1968) convenience sample from Boston
2	Loehlin and Nichols (1976) high-achieving high school students
3	Matheny et al. (1976) sample from Louisville, KY
4	Plomin et al. (1976) members of mothers of twins' organizations
5	Scarr and Carter-Saltzman (1979) convenience sample, Philadelphia
6	Rose et al. (1988) Finnish Twin Cohort Study (FTCS)
7	Kaprio et al. (1990) FTCS
8	Morris-Yates et al. (1990) Australian Twin Registry
9	Kendler et al. (1993) Virginia Twin Registry (VTR)
10	Kendler et al. (1994) VTR
11	Hettema et al. (1995) VTR
12	Allison et al. (1996) MZs raised apart from Finland, Japan, & US
13	LaBuda et al. (1997) patients at drug abuse & alcohol treatment center
14	Bulik (1998) white females from VTR
15	Kendler and Gardner (1998) VTR
16	Klump et al. (2000) Minnesota Twin Family Study (MTFS)
17	Xian et al. (2000) Vietnam Era Twin Registry
18	Borkenau et al. (2002) Convenience sample, Germany
19	Cronk et al. (2002) female population-based sample, Missouri
20	Horwitz et al. (2003) Nat'l Long. Study of Adolescent Health
21	Penninkilampi-Kerola and Moilanen (2005) FTCS
22	Derks et al. (2005) Netherlands Twin Register
23	Eriksson et al. (2006) Swedish Young Male Twins Study
24	Gunderson et al. (2006) Kaiser Permanente Twin Registry
25	Conley et al. (2013) Nat'l Long. Study of Adolescent Health, Swedish Twin Registry, MTFS

Table 2
Measures of environmental similarity used in studies of the equal environments assumption.

Co-dependence[21]
Emotional closeness[15 (in youth),13 (in adulthood)]
Frequency of contact in adulthood[6,7,13,14,23]
Frequency of contact in youth[2,8,12,14,15,18,19,20]
Model-estimated similarity in environment[22]
Similarity of treatment by parents or by others in youth[2,8,10,14,15,18]
Perception of zygosity[1,5,9,10,19,17,24,25]
Similarity of appearance[3,4,5,11,14,16,20]

Numbers in superscript refer to the studies enumerated in Table 1.

confounded estimates of heritability. Also notable is that few studies examined measures of environmental similarity from more than one of the categories listed in Table 2.

Table 3 shows the outcomes for which the EEA has been tested among the studies listed in Table 1. Most tests of the EEA have involved psychological traits, not sociological outcomes.

Table 4 indicates the methodology used in each study as well as the number of EEA violations found among relationships tested, when this could be ascertained. Violations of the EEA are defined as statistically significant relationships between measures of environmental similarity and outcome similarity between co-twins. I tallied these numbers for purposes of comparison, as they were not typically reported as such by the authors of the studies themselves.

Overall, only about 11% of findings from all twenty-five studies were significant, which is only a bit more than we would expect due to chance. Sample sizes were typically in the hundreds or thousands, so the power to detect sizeable effects is relatively high. The results of previous research suggest that the EEA is generally sound. Also, there are no patterns in the data suggesting either that particular aspects of the environment might matter more than others or that particular outcomes might be more affected than others.

Nearly all of the evidence to date is supportive of the EEA. However, results could be biased if the measurement of environmental similarity has been low in validity and/or low in reliability (Richardson and Norgate, 2005). Indeed, my reanalysis of L&N's data showed that improvements in the validity and reliability of measurement can have a fairly substantial impact on results. According to proponents of twin studies, these issues of reliability and validity are moot since heritability estimates from twin studies have been corroborated by other methods that do not rely on the equal environments assumption, namely adoption studies and studies of twins reared apart.

6. Studies of adopted children and studies of twins reared apart

If the EEA was consistently violated, then we might expect heritability estimates from twin studies to be higher than corresponding estimates based on other techniques that do not rely on this assumption. In fact, the results of adoption studies

Table 3
Outcomes variables in recent evaluations of the EEA.

Category	Outcomes & study Ref. numbers
Physiological traits	Height[3,25]
	Weight & body mass index[5,12,20,25]
	Other anthropometric & biometric measures[5,24]
Psychological traits	Agreeableness[2,15,18]
	Conscientiousness[18]
	Extroversion/sociability[3,4,2,5,6,7,18]
	Neuroticism[6,7,8,18]
	Openness to experience[18]
	Other personality measures[1,3,4]
	Anxiety disorders[1,8,9,10,11,15,17,19]
	Attention deficit disorder[19,25]
	Depression/well-being/self-esteem[5,2,8,9,10,11,15,17,20,25]
Health behaviors	Alcohol consumption & dependence[5,6,8,7,9,11,13,15,20,17,21]
	Diet[24]
	Eating disorders[9,11,15,16]
	Illicit drug dependence[13,17]
	Physical activity[23]
	Smoking[15,17]
Academic performance	High school GPA[25]
	Tests of academic achievement[2]
	Tests of cognitive skills[3,5,22]
Other traits	Anti-social behavior (aggression,delinquency, etc.)[2,19,22,25]
	Vocational interests[2]

Numbers in superscript refer to the studies enumerated in Table 1.

Table 4
Recent studies: methods and findings.

Study	Method	Violations
1	Perceived & actual zygosity	N/A
2	Correlations within zygosity	64 of 574
3	Rank-order correlations within zygosity	2 of 44
4	Correlations within zygosity	2 out of 16
5	Regressions	0 of 4
6	Regressions	2 of 2
7	Correlations	2 of 3
8	Partial correlations	1 of 5
9	SEM on contingency tables	0 of 5
10	SEM on contingency tables	0 of 4
11	SEM on contingency tables	1 of 5
12	Correlation	0 of 1
13	Logistic regressions	2 of 6
14	Logistic regressions	1 out 12
15	Logistic regressions	2 of 48
16	ANOVA, MANOVA, correlations	0 of 4
17	Multigroup SEM	0 of 4
18	Regressions	3 of 30
19	Multigroup SEM	0 of 4
20	Regressions	2 or 3 of 6
21	SEM on contingency tables	N/A
22	SEM	0 of 2
23	SEM on contingency tables	0 of 4
24	Correlations by correct & incorrect zygosity	1 of 9
25	Estimated heritabilities	0 of 5

and studies of twins reared apart are generally consistent with the results of classical twin studies. For example, in the Texas Adoption Study, the personalities of seventeen year olds who had been adopted within several days of birth were correlated with the personalities of their biological mothers, but uncorrelated with the personalities of their adoptive mothers (Rowe, 1995, p.69). This result is consistent with the results of classic twin studies in revealing substantial heritability and a negligible effect of shared (family) environment on personality.

But studies of adopted children are not unassailable. Since adoptive parents have typically undergone a screening process, the variation in adoptive environments is likely to be low. Also, adoption agencies may assign children to adoptive families in a non-random way that induces correlations between the traits of the children and traits of their biological mothers. For these reasons, some critics have suggested that the estimates of genetic influence from adoption studies may be biased Conley et al. (2013).

What about the results of studies of twins reared apart? There are only a few studies of twins reared apart, since data on such twins are scarce. If the results of classic twin studies were biased due to violations of the EEA, and results of studies of twins reared apart were unbiased, we would expect the heritability estimates to vary accordingly. Instead, the results of studies of twins reared apart are essentially consistent with the results of classic twin studies (Rowe, 1995, p. 72). The most prominent study of twins reared apart in the US revealed moderate to high correlations on a variety of characteristics (Bouchard et al., 1990), such as IQ and personality.

Nonetheless, the results of Bouchard et al. (1990) may be compromised by selection bias. Unsurprisingly, the data on twins reared apart in that study were not a representative sample of a known population. Twins were recruited through advertisements, and many subjects were reared in separate households but had met prior to participating in the study. Joseph (2010) suggested that the twins who elected to participate in the study were likely to be more similar to one another than twins who chose not to participate.

In general, the evidence supports the EEA, but there are reasons to think that heritability estimates from all three types of behavior genetic studies – adoption, twins reared apart and the classic twin study – *may* be biased upward.

7. Present study

In this study, I evaluate the EEA in a more comprehensive manner than any previous study has done. I measure environmental similarity in a more complete way by incorporating a greater variety of measures, and I estimate the reliability of these measures. I examine a wider range of outcomes than any previous study has done, which permits an assessment of whether bias varies across types of outcomes. It is possible that bias is more substantial for sociological outcomes than for physiological or psychological outcomes. This is because co-twin similarity on certain sociological outcomes may result in part from twins being in the same place at the same time in their lives, and MZ co-twins will be more likely than DZ co-twins to synchronize their lives since they identify with each other so strongly. Religious service attendance and education are two examples. We may observe greater similarity in religious service attendance between MZ co-twins at least in part because MZ co-twins will be more likely to accompany each other to services. Likewise, MZ co-twins may be more similar to each on educational attainment than are DZ co-twins because MZ co-twins are more likely to go to the same college and attend the same courses.

8. Data

I analyzed data from the twin component of the Midlife Development in the United States (MIDUS) survey (Ryff et al., 2011).[6] The MIDUS twin sample is a nationally representative survey of English-speaking, non-institutionalized twins ages 25 to 74 in the United States conducted in 1995. Recruitment of twins was accomplished by random digit dialing an initial sample of 50,000 people and asking whether the respondent or anyone in their immediate family was a twin. Among twins identified in this manner who were deemed eligible to complete the survey, about 56% completed the initial interview by phone. Of those who completed the initial interview, 93% completed a self-administered questionnaire by mail, for an overall response rate of 55.2%(60% $\times$ 93%).[7]

The MIDUS data are suitable for these analyses as they contain a diverse array of measures relating to twins' environmental similarity and a large set of measures of physiological, psychological and sociological outcomes.

8.1. Outcome measures

I identified 32 outcomes in the MIDUS twins data for which there was evidence of genetic influence.[8] These 32 outcome measures were classified into seven categories: health and health-related behaviors, mental well-being, personality characteristics, physical attributes, self-efficacy, social and religious beliefs, and social class.

Table 6 contains descriptive statistics for the 32 outcomes. The *Items* column indicates the number of measures on which each outcome is based. Height is included in part as a test of whether it is reasonable to control for environmental similarity. It is widely accepted that height is largely heritable within developed countries and few would argue that heritability estimates for height are confounded by environmental similarity. If controlling for environmental similarity reduces estimates of heritability for height, then we might have doubts about whether it makes sense to control for environmental similarity in the first place. On the other hand, if controls for environmental similarity affect estimates of heritability for

[6] The MIDUS survey was conducted by the John D. and Catherine T. MacArthur Foundation Research Network on Successful Midlife Development (MIDMAC), an interdisciplinary research group.

[7] This response rate is approximate, since there is uncertainty in the classification of cases as ineligible.

[8] Twin studies providing heritability estimates for these outcomes are as follows: height (Silventoinen et al., 2003); waist-to-hip ratio (Nelson et al., 1999); body mass index and chronic health conditions (Johnson and Krueger, 2005); alcohol problems, smoking (ever smoked & number of cigarettes per day) (Kendler et al., 2011); the "Big Five" personality characteristics (Jang et al., 1996); depression (Schnittker, 2010); life satisfaction, psychological well-being, social well-being, positive affect (Kendler et al., 2011); negative affect (Baker et al., 1992); perceived constraints, personal mastery (Littvay et al., 2011); educational attainment (Branigan et al., 2013); income, net worth (Schnittker, 2008); self-employment (Nicolaou and Shane, 2010); age at menarche (Towne et al., 2005); number of biological children (Rodgers et al., 2001); prosocial obligation (Lewis and Bates, 2011); racial prejudice (Truett et al., 1992); religious service attendance, personal religiosity, exclusivist beliefs, biblical literalism and born-again religious commitment (Bradshaw and Ellison, 2008).

other outcomes, but not for height, then we can have more confidence that controlling for environmental similarity is a sensible strategy (Horwitz et al., 2003).

Also included among the outcomes are variables relating to self-efficacy and self-control. These measures may be of interest to sociologists, as self-efficacy may predict status attainment. Recently, several political scientists argued and found evidence for the idea that socio-political attitudes arrayed from liberal-to-conservative were partially rooted in biological differences between people (Alford et al., 2005). The heritability of religiosity has also been an active area of research, including a rare instance of behavior geneticists publishing in a sociology journal (Eaves et al., 2008).

8.2. Measures of environmental similarity

Previous research testing the equal environments assumption has typically measured environmental similarity in terms of treatment by parents and the amount of time twins have spent together. But psychological closeness may matter as much or more than physical proximity. Twins who see each other rarely may nonetheless remain emotionally close and may exert substantial influence on each other. With the MIDUS data, I am able to measure the quality of the twin relationship, as well as the quantity of time twins spend together. I constructed six scales of environmental similarity: (1) the similarity of their childhood environment, (2) the proportion of their life lives together, (3) the frequency of contact between co-twins (4) the level of psychological intimacy within the twinship (5) the extent of advice seeking and giving between co-twins and (6) one overall composite scale. The questions comprising the scales are described in Table 5. Each measure was equal to the mean of responses from both twins when they were available, or to the response of one twin if responses from the other twin were missing. (About 5% of twin pairs were missing responses from one twin on at least one measure of environmental similarity.)

Upon standardizing the scales of environmental similarity, I found that MZ twins were at least one-quarter of a standard deviation higher on all measures of environmental similarity than were DZ twins. Environmental similarity diverged the most between MZ and DZ twins with respect to childhood treatment, and the least with respect to current contact.

8.3. Measuring zygosity

The measure of zygosity in MIDUS is based on questions asked of the twins about their level of physical similarity and how often people confused them when they were growing up. When validated by genetic measurement, these methods have been shown to be valid in over 90% of cases (Kasriel and Eaves, 1976).[9] Based on twins' answers to a set of survey questions, MIDUS investigators were able to classify over 98% of same-sex twins as MZ or DZ.

8.4. Analytic samples

It was not possible to conduct analyses for all outcomes on the same sample without excluding two-thirds of the cases. Thus, analysis for each outcome was conducted using all non-missing cases on that outcome. Cases where information from both twins was missing on any measure of environmental similarity were excluded. I also excluded cases when information was missing on one twin, when zygosity was undetermined, and in cases where the sample contained multiple pairs of twins from the same family.[10] These exclusions reduced the twin sample by 10% to 20% depending on the outcome. Sample sizes ranged from 196 MZ pairs and 153 DZ pairs for the analysis of job autonomy to 340 MZ and 305 DZ pairs for the analysis of depression.

9. Methods

For each of the 32 outcomes, the analysis proceeded in four stages. In the first stage, I estimated heritability for the outcome without controlling for environmental similarity. For outcomes that had more than two categories (27 out of 32), I estimated heritability using Defries-Fulker (DF) regression. For dichotomous outcomes (5 out of 32), I estimated heritability using a bivariate probit regression technique developed for this purpose by Kohler and Rodgers (1999).[11] I focus on the standard DF regression technique below and refer readers to Kohler and Rodgers (1999) for more information about the bivariate probit model.

The Defries-Fulker (DF) regression model provides unbiased estimates of heritability (see Smith and Hatemi, 2012 for details) and is illustrated in Eq. (3) below.

$$Y_1 = b_0 + b_{Y_2} Y_2 + b_h h + b_{hY_2} hY_2 + b_a age + b_g gender + e \tag{3}$$

[9] In the National Longitudinal Study of Adolescent Health, zygosity was initially measured via survey methods and later validated by genetic sequencing. Survey methods permitted zygosity classification for 93.4% of twins. Of these, 5.8% were misclassified.

[10] There were thirteen families who contributed two or more pairs of same-sex twins to the data. In order to maintain independent observations, I randomly chose a single twin pair from each of these families, and discarded the rest. This deletion resulted in the elimination of fifteen pairs of twins, or about 2.2% of the total sample of same-sex twins.

[11] This procedure assumes that the categorical outcome measures are manifestations of normally distributed latent variables.

Table 5
Environmental similarity scales.

Scales and component measures	Mean	Std	Min	Max
Similarity of childhood				
When you were children, how often did you and your twin have the same playmates?[a]	1.8	.79	1	4
How often were you in the same classroom in school?[a]	2.3	1.1	1	4
How often did you dress alike?[a]	2.3	1	1	4
Did you ever share the same bedroom with your twin during the time you were growing up? (1 = yes, 0 = no)	.98	.13	0	1
For how many years did the two of you have the same bedroom?	16	4.6	1	70
Proportion of life lived together				
Are you currently living with your twin? (1 = yes, 0 = no)	.032	.18	0	1
In your lifetime, how many total years have the two of you lived together?	19	3.3	0	50
Age of twins	45	12	25	74
Current contact				
How frequently do you and your twin see each other? (9 = live together, 8 = several times a day, 7 = once a day, 6 = several times a week, 5 = once a week, 4 = 2 or 3x a month, 3 = 1x a month, 2= ⩽ 1x month, 1 = never or hardly ever)	3.8	2.2	1	9
Psychological intimacy				
How much does your twin understand the way you feel about things?[b]	1.4	.72	1	4
How much can you rely on (him/her) for help if you have a serious problem?[b]	1.3	.64	1	4
How much does your twin really care about you?[b]	1.1	.36	1	4
How much can you open up to your twin if you need to talk about worries?[b]	1.3	.72	1	4
Advice giving and receiving				
When you have a personal or practical problem, how much of the time do you turn to your twin for advice or help?[a]	3.1	1.3	1	5
When your twin has a personal or practical problem, how often does (he/she) turn to you for advice or help?[a]	3.2	1.2	1	5

[a] 4 = always, 3 = most of the time, 2 = sometimes, 1 = never.
[b] 4 = a lot, 3 = some, 2 = a little, 1 = not at all.

In a DF regression model, the value of the trait for one of the twins (Y_1) is regressed on: (1) the value of the trait for the other twin (Y_2), (2) a variable h equal to 0.5 for DZ twins and 1 for MZ twins, and (3) the interaction between Y_2 and h. The coefficient for the interaction term (b_{hY_2}) is an estimate of heritability, and the coefficient for the twin's trait (b_{Y_2}) is an estimate of shared (family) environmental influence. The model also includes controls for age and gender.

In the second stage of analysis, I tested the significance of the effects of each scale of environmental similarity individually in separate D-F regressions. I did this by adding terms $b_S S + b_{SY_2} SY_2$ to the DF regression model, where S is the scale of environmental similarity and SY_2 is an interaction between the similarity scale and the twin's trait.[12] The b_{SY_2} term is an estimate of the effect of environmental similarity on twin similarity. Scales of environmental similarity for which this interaction was significant were identified and saved for the next step.

In the third stage, I estimated a DF regression that included the 'main effects' and interaction terms for environmental similarity scales that were significant in the second stage. (When no similarity scale was significant in the first stage, I included the scale with the highest t-value in the second stage.) I then tested the significance of the difference between the estimate of heritability with and without controlling for environmental similarity using a Wald test.[13] Significance levels for these tests were calculated according to procedures discussed in the section on family-wise error rate below.

As a sensitivity check, I sought to replicate the bulk of these analyses using data from the second wave of MIDUS, collected between 2004 and 2006. This was possible for 26 out of the 32 original outcomes.

9.1. Controlling the family-wise error rate

The risk of falsely rejecting one or more null hypotheses is inflated here not only because 32 hypotheses are being tested, but also because the similarity scales were pretested for significance in the first stage of analysis described above. For this reason, conventional standards of significance were not appropriate here. On the other hand, procedures that adjust significance thresholds based only on the number of tests (i.e. Bonferroni) may be too conservative as they assume that hypothesis tests are independent. Type I error inflation is not likely to be as severe when hypothesis tests are correlated, as they almost certainly are here.

In order to achieve a balance between Type I and Type II error, I conducted simulations to estimate the distribution of results one would obtain if environmental similarity did not confound heritability for any outcome – a 'global null hypothesis.'[14] Simulations involved three steps. First, I used the correlation matrices from the MZ and DZ twins in the existing dataset to simulate a set of 74 correlated random normal variables (32 outcomes and 5 environmental similarity scales for each twin). Secondly, many of these variables were then transformed so that their distributions would roughly match the distributions found in the MIDUS data – dichotomous, left-skewed and right-skewed. Third, the main DF regression analyses were conducted.

[12] In the bivariate probit regressions used for dichotomous outcomes, the measure of environmental similarity is included as a predictor of the (latent) correlation between traits of the co-twins.
[13] I also included BIC tests for reference.
[14] I also report results and arrive at the same conclusions using the more conservative Bonferroni standard.

Table 6
Descriptive statistics for outcomes.

Outcome	Mean	Std	Min	Max
Health & health behaviors				
Chronic conditions	2.2	2.3	0	21
Alcohol problems	.06	.24	0	1
Respondent ever smoked	.75	.43	0	1
Cigarettes/day at peak	11	16	0	97
Mental well-being				
Depression	.12	.32	0	1
Life satisfaction	8	1.5	0	10
Psychological well-being	101	14	48	126
Social well-being	65	13	24	98
Positive affect	3.5	.71	1	5
Negative affect	1.5	.62	1	5
Personality				
Agreeableness	3.5	.46	1.2	4
Conscientiousness	3.4	.44	1.8	4
extraversion	3.2	.56	1.4	4
Neuroticism	2.2	.67	1	4
Openness to experience	3	.53	1	4
Physical attributes				
Height (inches)	67	4.1	55	83
Waist-to-hip ratio	.87	.096	.67	1.4
Body Mass Index	26	5.1	15	57
Reproductive behavior				
Age at menarche (women only)	13	1.6	9	21
Number of biological children	1.9	1.4	0	5
Self-efficacy				
Perceived constraints	2.6	1.2	1	6.6
Personal mastery	5.9	.98	1	7
Social and religious beliefs				
Prosocial obligation	77	18	7	110
Racial prejudice[c]	6.1	3.1	1	13
Religious attendance	2.9	1.4	1	5
Personal religiosity	-.028	.79	−1.1	2.2
Exclusivist beliefs	4.8	2.4	1	9
Biblical literalism[a]	3.6	1.4	1	5
Born-again religious commitment	.52	.5	0	1
Social class				
Education in years	13	2.5	4	20
Income	26	25	0	125
Net worth	103	194	−175	1000
Self-employment	.33	.47	0	1

[a] Only asked of self-identified Christians.

The difference between estimates of heritability with and without controlling for environmental similarity, as well as the significance of that difference, were recorded. These three steps were replicated 10,000 times, and minimum *p*-values were recorded. The fifth percentile of the distribution of simulated *p*-values was approximately 0.0024; this is the value I use as a cut-off for statistical significance in the analysis below. (Note that the corresponding Bonferroni and Šidák adjusted *p*-value thresholds for 32 tests are approximately 0.001602 and 0.001563, respectively.) We can also learn from the simulation that, when environmental similarity has no confounding effects, using the conventional threshold of significance, there is a 62% chance of falsely rejecting at least one hypothesis but only about a 9% chance of falsely rejecting three more null hypotheses. The results of the simulation will be useful in interpreting the results of the data analysis.

9.2. Do distributions of outcomes vary by zygosity?

It is customary in twin studies to ensure that distributions of outcomes are similar across zygosity. I tested for significant differences in the means and variances of each outcome variable between MZ twins and DZ twins using *t*-tests and Levene's tests (Levene, 1960), respectively. Trait variances did not differ significantly by zygosity; however, there were some small, significant differences in trait means by zygosity. MZ twins reported significantly fewer chronic conditions, more exercise, less smoking, better mental health, more years of education, and a higher net worth than did DZ twins. These differences are small and are unlikely to be consequential in the analysis.

9.3. Reliability of measurement

Richardson and Norgate (2005) argued that the reliabilities for measures of environmental similarity are between 0.3 and 0.5. Reliabilities this low could bias effects of environmental similarity substantially downward. In the present study, it is possible to estimate the reliability for the scales of environmental similarity that measure facts about their lives together. Co-twin reports of factual information about their joint experiences (i.e how often they dressed alike when young, etc.) can be considered parallel measures since co-twins are reporting on the same thing and will have the same error variance (Alwin, 2007, pp. 40–41). The reliability of parallel measures is equal to the correlation between the two measures (Alwin, 2007, pp. 40–41). Measured in this way, the reliabilities of *individual* twin reports of similarity of childhood, proportion of life lived together and current contact are approximately 0.64,0.95, and 0.92, respectively. Since the scales used in the analysis are averages of individual co-twin reports, effective reliabilities for scales used in the analysis are higher, specifically 0.78 (childhood), 0.97 (proportion of life) and 0.96 (current contact).[15]

Reliability cannot be estimated as precisely for the other two scales of environmental similarity – psychological intimacy and advice – since the items that comprise these scales measure twins' own opinions and are thus not parallel measures. However, in the absence of nonrandom error, a lower bound estimate of reliability can be estimated with Cronbach's alpha (Alwin, 2007, p. 53). In this way, the reliabilities for the scales of psychological intimacy and advice giving and receiving are estimated to be no lower than 0.80 and 0.87, respectively. Overall, then, reliabilities of environmental similarity scales are generally rather high.

Although it is generally possible to adjust estimates for unreliability in regression using errors-in-variables regression, it was not possible to estimate Defries-Fulker models using this technique.[16] In any case, it would be highly speculative to estimate models using the lower-bound estimates of reliability for psychological intimacy and advice giving. Thus, the results presented below are not adjusted for unreliability. To gauge the extent to which unreliability might reduce the confounding effects of environmental similarity, I estimated errors-in-variables regressions of the absolute differences between co-twins in each of the 32 outcomes on a dummy variable for monozygotic twins with and without controlling for the five measures of environmental similarity. Reliabilities for psychological intimacy and advice giving and receiving were set at their estimated lower bounds and reliabilities for the other scales were set at the corresponding figures mentioned above. I also estimated reliability for zygosity at 0.90 by correlating co-twin reports on the questions used to determine by zygosity. On average across 32 outcomes, adjusting for unreliability increased the extent of apparent confounding by 8%. This is an upper-bound estimate, since the reliabilities of two environmental similarity scales were set at their lower bounds.

10. Results

The main results are displayed in Figs. 1 and 2. Outcomes are sorted according to the amount of reduction in heritability after accounting for environmental similarity. Fig. 1 shows the sixteen outcomes for which heritability fell the most after inclusion of controls for environmental similarity. Fig. 2 shows heritability estimates for the remaining sixteen outcomes.

Heritability estimates without controls are indicated by triangles. Heritability estimates with controls for environmental similarity included are indicated by circles. Solid triangles and circles represent heritability estimates that are significantly different from zero. Empty triangles and circles represent heritability estimates that do not differ significantly from zero.

Recall that the threshold for statistical significance for a change in heritability was determined through simulation. The fifth percentile of the minimum *p*-values generated by the simulation was approximately equal to 0.0024. The reduction in heritability met this threshold only for neuroticism.[17] Heritability for neuroticism dropped 31% from 0.40 to 0.28. This result is indicated with a dark arrow on the chart.

Although only one reduction in heritability was significant, the average magnitude of the reductions was more than what we would expect if there was no environmental confounding at all. In simulations of models with no confounding, heritability reductions averaged about 7%. In the real data, heritability reductions averaged about 14%, excluding the reduction for neuroticism. Heritability was reduced to some extent for 28 out of 32 outcomes. In the simulation, heritability was reduced for only 17 out of 32 outcomes on average. These results suggest that confounding is not negligible, but modest enough to go undetected for many outcomes.

Confounding does not appear to be patterned across outcomes. For instance, heritability is not reduced more for sociological outcomes like education and religious attendance than it is for psychological outcomes like extraversion and conscientiousness. In addition, estimates of confounding vary dramatically across outcomes that are similar, i.e between psychological well-being, depression and life satisfaction; and between education, income and net worth. The lack of discernible patterns here suggests that variation in apparent confounding across outcomes reflects chance fluctuation.

[15] The reliability of the average of two parallel measures is equal to $\frac{r}{r+.5\times(1-r)}$ where r is the reliability of each individual measure. This formula was derived algebraically and confirmed with simulations. A proof is available from the author.

[16] In order to estimate errors-in-variables regressions, all estimates of reliability must exceed the R^2 for models of each independent variable regressed on all other variables in that model. The Defries-Fulker model contains interactions which are strongly predicted by constituent variables, making it virtually impossible to run errors-in-variables regression in this case.

[17] Other procedures for controlling the family-wise error rate – Šidák and Bonferroni-Holm – also yielded only one significant reduction in heritability.

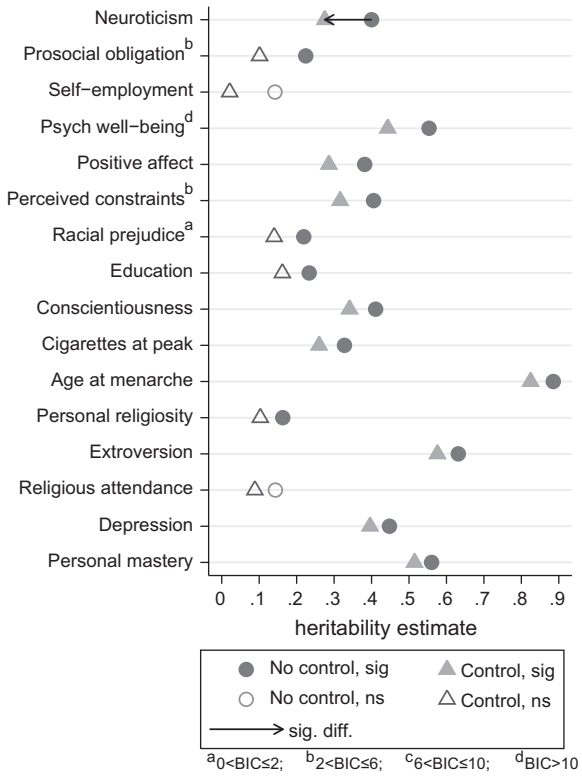

Fig. 1. Estimates of heritability, with and without controls.

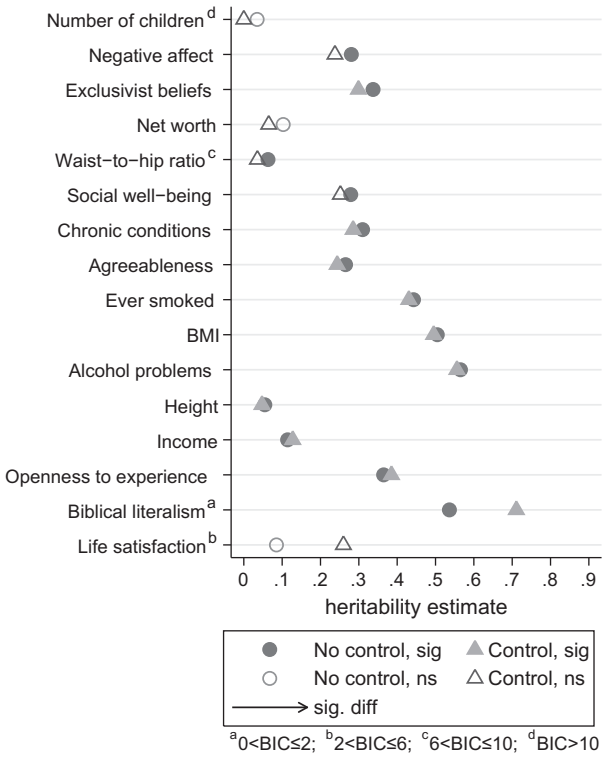

Fig. 2. Estimates of heritability, with and without controls, continued.

Further evidence that differences in apparent confounding across outcomes is due to random processes comes from comparisons with results from wave two. Among the 26 outcomes that exist at both waves, heritability reduction between waves correlates at about 0.44, and when one outlier (life satisfaction) is removed, the correlation falls to 0.20. Thus, consistency of the results across waves is rather weak. Again, this suggests that apparent differences in confounding across outcomes are generated mostly by random error.

11. Conclusion

Controlling for environmental similarity reduced heritability significantly for only one out of the 32 outcomes examined: neuroticism. It is unclear however whether the EEA is particularly problematic with respect to neuroticism. Heritability reduction for neuroticism in the second wave of MIDUS was far from significant, and previous research is mixed on the question. On the one hand, Rose et al. (1988) found that co-twin interaction was associated with similarity in neuroticism scores, and Kaprio et al. (1990) found that twins who had less contact with each other became less similar with respect to neuroticism. On the other hand, Morris-Yates et al. (1990) and Borkenau et al. (2002) found no evidence that co-twin similarity in neuroticism was related to environmental similarity. More broadly, there is little evidence that close contact engenders similarity on neuroticism. For these reasons, it is premature to conclude that the EEA is particularly problematic in studies of neuroticism.

In considering the results as a whole, the near lack of any significant results seem to suggest that the EEA is valid in most cases, but the evidence does not favor a 'global null' hypothesis. In the main analysis, heritability estimates for 19 out of 32 outcomes were reduced by 10% or more when controlling for environmental similarity. This result is notable given that none of the 10,000 simulations of these models under the global null hypothesis yielded more than 17 outcomes with reductions greater than 10%. If the global null hypothesis were true, and the simulations provide sufficient guidance, the results that were obtained would have less than .01% chance of occurring.

The reanalysis of L&N was also inconsistent with the idea that the EEA is valid for all outcomes. Among the correlations between 82 measures of outcome similarity and a composite measure of twin-reported environmental similarity, 33 (40%) were above 0.10. If all correlations were actually equal to zero, we would expect fewer than 2% of them to be greater than 0.10 in samples of the size analyzed by L&N. Taken as a whole, the evidence suggests that heritability for many if not all outcomes is confounded to a degree small enough to pass undetected in many studies. My assessment of reliability suggests that these confounding effects would be somewhat larger if measures of environmental similarity were perfectly reliable.

My results also suggest that confounding does not vary much across outcomes. This conclusion is based in part on the fact that results were inconsistent between waves one and two of MIDUS. Also, there was no evidence that heritability estimates for sociological outcomes are more affected than heritability estimates for psychological outcomes. Perhaps this is not surprising since, logically speaking, much of the genetic impact on sociological variables is mediated by psychological characteristics. In any event, much of the variation across outcomes in apparent confounding is likely due to random error around a nonzero true value.

Proponents of twin studies may disagree with this conclusion on the grounds that controlling for environmental similarity controls away effects that are properly attributed to genes. Proponents may likewise remind us that heritability includes the effects of environments chosen on the basis of genetic predispositions. In other words, environmental similarity is not entirely exogenous. For example, the genetic effect on neuroticism is perhaps partly due to the fact that people who are predisposed to experience negative affect will spend more time with others who are similarly predisposed. Hence, controlling for the time that twins spend together may control away part of the true genetic effect on neuroticism.

Endogeneity may be less of a problem in studies that test the EEA by examining outcome similarity among twins who misperceive their zygosity, i.e. MZ twins who believe they are DZ and DZ twins who believe they are MZ. This is because perceptions of zygosity most likely trace back to judgments made by doctors and parents shortly after the birth of the twins (Conley et al., 2013). Hence, it is notable that the most thorough examination of twins of mistaken zygosity concluded that the EEA was essentially sound (Conley et al., 2013). Conley et al. (2013) found that correlations on BMI, height, ADHD, depression and GPA were no lower – and in some cases were actually higher – between MZ co-twins who believed they were dizygotic than between twins who accurately perceived their zygosity.

Nonetheless, twin study skeptics may argue that the environments created by MZ twins are sui generis and will produce effects that cannot be generalized to non-twins. The idiosyncratic bonds produced by a pair of genetic clones may not be fully captured by perceptions nor by any of the measures of environmental similarity used in the present study. Differences in treatment experienced by DZ co-twins relative to MZ co-twins may be subtle enough to evade measurement but important enough to bias heritability estimates. Twins may also be hard-pressed to acknowledge such differences to themselves or to others. Finally, estimates of heritability could be inflated because MZ co-twins may be content in their similarity while DZ twins seek to differentiate themselves.

These criticisms notwithstanding, the evidence presented here suggests a middle ground between extreme positions on either side of the debate. All things considered, it seems unlikely that the EEA is strictly valid, but it also seems likely that violations of the EEA are relatively modest. In light of this evidence, perhaps more social scientists should adopt the perspective of Freese (2008), who argued that the evidence for (nearly) ubiquitous effects of genes on individual-level outcomes was

compelling, but that this conclusion did not necessarily undermine the role of sociology in providing explanations for social behavior.

Even when heritabilities are high, sociology remains vital because genetic causes of behavior are necessarily distal and contingent. Establishing a link between genes and a behavioral outcome does not by itself count as an explanation of variation in that outcome. Heritability estimates for behavioral outcomes may reflect a heterogeneous set of influences that is almost invariably contingent on the environment in some way (Jencks, 1980). Consider the heritability for depression – estimated at 0.40 in this study. This heritability estimate could arise from the actions of genes that code for biochemical processes involved in producing negative affect. But the heritability estimate could also arise from the effects of a gene-environment interaction. For instance, suppose in one case, a child is raised by a caregiver whose parenting style is congruent with the child's genetic tendencies, while in another case there is a mismatch between parenting style and genetic predisposition. In this hypothetical scenario, it is conceivable that the first child would be less susceptible to depression than the second. In this scenario, what would appear as a genetic effect in a twin study would actually be as much the result of the environment as of genes.

This is only one hypothetical example of how the path from genes to outcomes is "blocked" by what Freese (2008) called a "phenotypic bottleneck" in that genes cannot logically affect actions outside of a social context. Social action necessarily involves aspects of the situation and the larger social world and hence requires sociological as well as psychological and biological explanation. Nonetheless, the results presented here suggest that genes do play an important role in affecting many outcomes of interest to sociologists.

Appendix A

See Table A.1.

Table A.1
Components of scales.

Construct	Questionnaire items
Height (inches)	How tall are you?
Waist-to-hip ratio[a]	What is your waist size – that is, how many inches around is your waist?; What is your hip size – that is, how many inches do your hips measure at the widest point? Measure at the widest point between your waist and your thighs. [Respondents were provided with a tape measure]
Body Mass Index[a]	How much do you currently weigh?; How tall are you?
Chronic conditions[a]	In the past 12 months, have you experienced or been treated for any of the following: asthma, bronchitis, emphysema, tuberculosis, other lung problems, bone or joint diseases, sciatica, lumbago, recur backache, persistent skin trouble, thyroid disease, hay fever, recurring stomach trouble, urinary or bladder problems, being constipated, gall bladder trouble, persistent foot trouble, varicose veins requiring treatment, AIDS or HIV infection, autoimmune disorders, trouble with your gums or mouth, persistent trouble with your teeth
Alcohol problems[d]	Were you ever, during the past 12 months, under the effects of alcohol or feeling its after-effects in a situation which increased your chances of getting hurt - such as when driving a car or boat, or using knives or guns or machinery?; Did you ever, during the past 12 months, have any emotional or psychological problems from using alcohol – such as feeling depressed, being suspicious of people, or having strange ideas?; Did you ever, during the past 12 months, have such a strong desire or urge to use alcohol that you could not resist it or could not think of anything else?; Did you have a period of a month or more during the past 12 months when you spent a great deal of time using alcohol or getting over its effects?; Did you ever, during the past 12 months, find that you had to use more alcohol than usual to get the same effect or that the same amount had less effect on you than before?
Respondent ever smoked[d]	Have you ever smoked cigarettes regularly – that is, at least a few cigarettes every day?
Cigarettes/day at peak	On average, about how many cigarettes did you smoke per day during the one year in your life when you smoked most heavily?
Agreeableness [a,c]	Helpful; Warm; Caring; Softhearted; Sympathetic
Conscientiousness [a,c]	Organized; Responsible; Hardworking; Careless
Extraversion[a,c]	Outgoing; Friendly; Lively; Active; Talkative
Neuroticism[a,c]	Moody; Worrying; Nervous; Calm
Openness to experience[a,c]	Creative; Imaginative; Intelligent; Curious; Broad-minded; Sophisticated; Adventurous
Depression[d]	During the past 12 months, was there ever a time when you felt sad, blue or depressed for two weeks or more in a row?; [If yes] During those two weeks, did you: feel more tired out or low on energy than is usual, lose your appetite, have more trouble falling asleep than usual, have a lot more trouble concentrating than usual, feel down on yourself, no good, or worthless, think a lot about death
Life satisfaction	Rate your life overall these days on a scale from 0 to 10 where 0 = worst possible life overall and 10 = the best possible life overall
Psychological well-being	Self-acceptance: I like most parts of my personality; When I look at the story of my life, I am pleased with how things have turned out so far; In many ways I feel disappointed about my achievements in life; Personal growth: For me, life has been a continuous process of learning, changing and growth; I think it is important to have new experiences that challenge how I think about myself and the world; I gave up trying to make big improvements or changes in my life a long time ago; Purpose in life: Some people wander aimlessly through life, but I am not one of them; I live life one day at a time and do not really think about the future; I sometimes feel as if I have done all there is to do in life; Environmental

Table A.1 (*continued*)

Construct	Questionnaire items
	mastery: The demands of everyday life often get me down; In general, I feel I am in charge of the situation in which I live; I am good at managing the responsibilities of daily life; Autonomy: I tend to be influenced by others with strong opinions; I have confidence in my own opinions, even if they are different from the way most other people think; I judge myself by what I think is important, not by the values of what others think is important; Positive relations with others: Maintaining close relationships has been difficult and frustrating for me; People would describe me as a giving person, willing to share my time with others; I have not experienced many warm and trusting relationships with others.
Social well-being[e]	Social acceptance: People who do you a favor expect nothing in return; People do not care about other people's problems; I believe that people are kind; Social growth: The world is becoming a better place for everyone; Society has stopped making progress; Society is not improving for people like me; Social contribution: I have something valuable to give to the world; My daily activities do not create anything worthwhile for my community; I have nothing important to contribute to society; Social coherence: The world is too complex for me; I cannot make sense of what's going on in the world; I find it easy to predict what will happen next in society; Social integration: I do not feel I belong to anything I had call a community; I feel close to other people in my community; My community is a source of comfort
Positive affect[a,f]	Cheerful, In good spirits, Extremely happy, Calm and peaceful, Satisfied, Full of life
Negative affect[a,f]	So sad nothing could cheer you up, Nervous, Restless or fidgety, Hopeless, That everything was an effort, Worthless
Perceived constraints[a,e]	There is little I can do to change the important things in my life; I often feel helpless in dealing with the problems of li.e.; Other people determine most of what I can and cannot do.; What happens in my life is often beyond my control.; There are many things that interfere with what I want to do.; I have little control over the things that happen to me.; There is really no way I can solve the problems I have.; I sometimes feel I am being pushed around in my life.
Personal mastery[a,e]	I can do just about anything I really set my mind to;When I really want to do something, I usually find a way to succeed at it.;Whether or not I am able to get what I want is in my own hands.; What happens to me in the future mostly depends on me.
Education in years	What is the highest grade of school or year of college you completed?
Income[a]	What was your own personal earnings income in the past 12 months, before taxes?
Net worth	Suppose you (and your spouse or partner) cashed in all your checking and savings accounts, stocks and bonds, real estate, sold your home, your vehicles, and all your valuable possessions. Then suppose you put that money toward paying off your mortgage and all your other loans, debts, and credit cards. Would you have any money left over after paying your debts or would you still owe money? How much would that be (that you had left over, or would owe)?
Self-employment[d]	Are you currently self-employed? Ten years ago, were you self-employed? Ten years from now, do you think you will be self-employed?
Age at menarche (women only)	How old were you when you had your first menstrual period?
Number of biological children	How many biological children do you have?
Prosocial obligation	How much obligation would you feel: to serve on a jury if called; to keep fully informed about national news and public issues; to testify in court about an accident you witnessed; to vote in local and national elections; to do more than most people would do on your kind of job; to work hard even if you did not like or respect your employer or supervisor; to cancel plans to visit friends if you were asked but not required; to work overtime; to pay more for your health care so that everyone had access to health care; to volunteer time or money to causes you support; to collect contributions for heart or cancer research if asked to do so; to vote for a law that would help others worse off than you but would increase your taxes (Answers for each ranged from 0 = none to 10 = great)
Racial prejudice[c]	How much do you prefer to be with people of the same ethnic group?; How important do you think it is for people who are from this ethnic group to marry other people who are also from this ethnic group?; How much do you prefer to be with other people who are the same race as yourself?; How important do you think it is for people who are in your racial group to marry other people who are the same race?
Religious attendance	How often do you usually attend religious or spiritual services? (more than once a week, about once a week, one to three times a month, less than once a month)
Personal religiosity[g]	How religious are you; How spiritual are you; How important is religion in your life; How important is spirituality in your life; How important is it for you–or would it be if you had children now – to send your children for religious or spiritual services or instruction?; How closely do you identify with being a member of your religious group?; When you have decisions to make in your daily life, how often do you ask yourself what your religious or spiritual beliefs suggest that you do?
Exclusivist beliefs[g]	Which of the following do you believe: that it is good to explore many different religious or spiritual teachings, or that one should stick to a particular faith?; How much do you prefer to be with people of the same religion as you?; How important do you think it is for people of your religion to marry other people who are the same religion? (very, somewhat, not very, not at all)
Biblical literalism[b]	Please tell me how much you agree or disagree with the following statement: The Bible is the actual Word of God and is to be taken literally, word for word.
Born-again religious commitment[d]	Have you been born again, that is, had a turning point in your life when you committed yourself to Jesus Christ?

Answer choices: all of the time, most of the time, some of the time, a little of the time, none of the time.

[a] Measure or scale was constructed by MIDUS investigators.

[b] Only asked of self-identified Christians.

[c] Please indicate how well each of the following describes you: a lot, some, a little, not at all.

[d] Answer choices: yes, no.

[e] Please indicate how strongly you agree or disagree with each of the following statements: agree strongly, agree somewhat, agree a little, do not know, disagree a little, disagree somewhat, disagree strongly.

[f] During the past 30 days, how much of the time did you feel: all of the time, most of the time, some of the time, a little of the time, none of the time.

[g] Answer choices: very, somewhat, not very, not at all.

References

Adkins, D.E., Vaisey, S., 2009. Toward a unified stratification theory: structure, genome, and status across human societies. Sociological Theory 27, 99–121.
Alford, J.R., Funk, C.L., Hibbing, J.R., 2005. Are political orientations genetically transmitted? American Political Science Review 99, 153–167.
Allison, D.B., Kaprio, J., Korkeila, M., Koskenvuo, M., Neale, M.C., Hayakawa, K., 1996. The heritability of body mass index among an international sample of monozygotic twins reared apart. International Journal of Obesity 20, 501–506.
Alwin, D.F., 2007. Margins of Error: A Study of Reliability in Survey Measurement. John Wiley & Sons, Hoboken, NJ.
Baker, L.A., Cesa, I.L., Gatz, M., Mellins, C., 1992. Genetic and environmental influences on positive and negative affect: support for a two-factor theory. Psychology and Aging 7, 158–163.
Benjamin, D.J., Cesarini, D., van der Loos, M.J.H.M., Dawes, C.T., Koellinger, P.D., Magnusson, P.K.E., Chabris, C.F., Conley, D., Laibson, D., Johannesson, M., Visscher, P.M., 2012. The genetic architecture of economic and political preferences. Proceedings of the National Academy of Sciences 109, 8026–8031.
Borkenau, P., Riemann, R., Angleitner, A., Spinath, F., 2002. Similarity of childhood experiences and personality resemblance in monozygostic and dizygotic twins: a test of the equal environments assumption. Personality and Individual Differences 33, 261–269.
Bouchard, T., Lykken, D., McGue, M., Segal, N., Tellegen, A., 1990. Sources of human psychological differences: the minnesota study of twins reared apart. Science 250, 223.
Bradshaw, M., Ellison, C., 2008. Do genetic factors influence religious life? Findings from a behavior genetic analysis of twin siblings. Journal for the Scientific Study of Religion 47, 529–544.
Branigan, A.R., McCallum, K.J., Freese, J., 2013. Variation in the heritability of educational attainment: An international meta-analysis. Northwestern University Institute for Policy Research Working Paper 13.
Bulik, C., 1998. Heritability of binge-eating and broadly defined bulimia nervosa. Biological Psychiatry 44, 1210–1218.
Chabris, C.F., Hebert, B.M., Benjamin, D.J., Beauchamp, J., Cesarini, D., van der Loos, M., Johannesson, M., Magnusson, P.K., Lichtenstein, P., Atwood, C.S., 2012. Most reported genetic associations with general intelligence are probably false positives. Psychological Science 23, 1314–1323.
Charney, E., 2012. Behavior genetics and postgenomics. Behavioral and Brain Sciences 35, 331–358.
Conley, D., Rauscher, E., Dawes, C., Magnusson, P.K., Siegal, M.L., 2013. Heritability and the equal environments assumption: evidence from multiple samples of misclassified twins. Behavior Genetics, 1–12.
Cronk, N.J., Slutske, W.S., Madden, P.A.F., Bucholz, K., Reich, W., Heath, A., 2002. Emotional and behavioral problems among female twins: an evaluation of the equal environments assumption. Journal of the American Academy of Child Adolescent Psychiatry 41, 829–837.
Davies, G., Tenesa, A., Payton, A., Yang, J., Harris, S.E., Liewald, D., Ke, X., Le Hellard, S., Christoforou, A., Luciano, M., McGhee, K., Lopez, L., Gow, A.J., Corley, J., Redmond, P., Fox, H.C., Haggarty, P., Whalley, L.J., McNeill, G., Goddard, M.E., Espeseth, T., Lundervold, A.J., Reinvang, I., Pickles, A., Steen, V.M., Ollier, W., Porteous, D.J., Horan, M., Starr, J.M., Pendleton, N., Visscher, P.M., Deary, I.J., 2011. Genome-wide association studies establish that human intelligence is highly heritable and polygenic. Molecular Psychiatry 16, 996–1005.
Derks, E.M., Dolan, C.V., Boomsma, D.I., 2005. A test of the equal environment assumption (EEA) in multivariate twin studies. Twin Research and Human Genetics 9, 403–411.
Eaves, L., Hatemi, P., Prom-Womley, E., 2008. Social and genetic influences on adolescent religious attitudes and practices. Social Forces 86, 1621–1646.
Eriksson, M., Rasmussen, F., Tynelius, P., 2006. Genetic factors in physical activity and the equal environment assumption – the Swedish young male twins study. Behavior Genetics 36, 238–247.
Freese, J., 2008. Genetics and the social science explanation of individual outcomes. American Journal of Sociology 114, S1–S35.
Gunderson, E., Ai-Lin, T., Selby, J., Caan, B., Mayer-Davis, E., Risch, N., 2006. Twins of mistaken zygosity (TOMZ): evidence for genetic contributions to dietary patterns and physiologic traits. Twin Research and Human Genetics 9, 540–549.
Guo, G., Roettger, M.E., Shih, J.C., 2007. Contributions of the DAT1 and DRD2 genes to serious and violent delinquency among adolescents and young adults. Human Genetics 121, 125–136.
Guo, G., Stearns, E., 2002. The social influences on the realization of genetic potential for intellectual development. Social Forces 80, 881–910.
Harris, J.R., 1999. The Nurture Assumption: Why Children Turn Out the Way They Do. Free Press, New York.
Hettema, J.M., Neale, M.C., Kendler, K.S., 1995. Physical similarity and the equal environments assumption in twin studies of psychiatric disorders. Behavior Genetics 25, 327–335.
Holmes, S.J., 1930. Nature vs. nurture in the development of the mind. The Scientific Monthly 31, 245–252.
Horwitz, A., Videon, T., Schmitz, M., Davis, D., 2003. Rethinking twins and environments: Possible social sources for presumed genetic influences in twin research. Journal of Health and Social Behavior 44, 111–129.
Jang, K.L., Livesley, W.J., Vemon, P.A., 1996. Heritability of the big five personality dimensions and their facets: a twin study. Journal of Personality 64, 577–592.
Jencks, C., 1980. Heredity, environment, and public policy reconsidered. American Sociological Review, 723–736.
Johnson, W., Krueger, R.F., 2005. Higher perceived life control decreases genetic variance in physical health: evidence from a national twin study. Journal of Personality and Social Psychology 88, 165.
Joseph, J., 2010. The genetics of political attitudes and behavior: claims and refutations. Ethical Human Psychology and Psychiatry 12, 200–217.
Kaprio, J., Koskenvuo, M., Rose, R.J., 1990. Change in cohabitation and intrapair similarity of monozygotic (MZ) cotwins for alcohol use, extraversion, and neuroticism. Behavior Genetics 20, 265–276.
Kasriel, J., Eaves, L., et al, 1976. The zygosity of twins: further evidence on the agreement between diagnosis by blood groups and written questionnaires. Journal of Biosocial Science 8, 263–266.
Kendler, K.S., Gardner, C.O., 1998. Twin studies of adult psychiatric and substance dependence disorders: are they biased by differences in the environmental experiences of monozygotic and dizygotic twins in childhood and adolescence? Psychological Medicine 28, 625–633.
Kendler, K.S., Myers, J.M., Maes, H.H., Keyes, C.L., 2011. The relationship between the genetic and environmental influences on common internalizing psychiatric disorders and mental well-being. Behavior Genetics 41, 641–650.
Kendler, K.S., Neale, M.C., Heath, A., Eaves, L., 1994. Parental treatment and the equal environment assumption in twin studies of psychiatric illness. Psychological Medicine 24, 579–590.
Kendler, K.S., Neale, M.C., Kessler, R., Heath, A., 1993. A test of the equal-environment assumption in twin studies of psychiatric illness. Behavior Genetics 23, 21–27.
Klump, K.L., Holly, A., Iacono, W.G., Mcgue, M., Willson, L.E., 2000. Physical similarity and twin resemblance for eating attitudes and behaviors: a test of the equal environments assumption. Behavior Genetics 30, 51–58.
Kohler, H.P., Rodgers, J.L., 1999. DF-like analyses of binary, ordered, and censored variables using probit and tobit approaches. Behavior Genetics 29, 221–232.
LaBuda, M., Svikis, D., Pickens, R., 1997. Twin closeness and co-twin risk for substance use disorders: assessing the impact of the equal environment assumption. Psychiatry Research 70, 155–164.
Levene, H., 1960. Robust tests for equality of variances. In: Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling, vol. 2, p. 278.
Lewis, G.J., Bates, T.C., 2011. A common heritable factor influences prosocial obligations across multiple domains. Biology Letters 7, 567–570.
Littvay, L., Weith, P.T., Dawes, C.T., 2011. Sense of control and voting: a genetically-driven relationship. Social Science Quarterly 92, 1236–1252.
Loehlin, J., Nichols, R., 1976. Heredity. Environment and Personality: A Study of 850 Sets of Twins. University of Texas Press, Austin.
Manolio, T., Collins, F., Cox, N., Goldstein, D., Hindorff, L., Hunter, D., McCarthy, M., Ramos, E., Cardon, L., Chakravarti, A., et al, 2009. Finding the missing heritability of complex diseases. Nature 461, 747–753.

Matheny, A.P., Wilson, R.S., Dolan, A.B., 1976. Relations between twins' similarity of appearance and behavioral similarity: testing an assumption. Behavior Genetics 6, 343–351.
Morris-Yates, A., Andrews, G., Howie, P., Henderson, S., 1990. Twins: a test of the equal environments assumption. Acta Psychiatry Scandinavia 81, 322–326.
Mowrer, E., 1954. Some factors in the affectional adjustment of twins. American Sociological Review 19, 468–471.
Nelson, T.L., Vogler, G.P., Pedersen, N.L., Miles, T.P., 1999. Genetic and environmental influences on waist-to-hip ratio and waist circumference in an older swedish twin population. International Journal of Obesity 23, 449–455.
Nicolaou, N., Shane, S., 2010. Entrepreneurship and occupational choice: genetic and environmental influences. Journal of Economic Behavior & Organization 76, 3–14.
Nielsen, F., 2006. Achievement and ascription in educational attainment: genetic and environmental influences on adolescent schooling. Social Forces 85, 193–216.
Nielsen, F., Roos, J., 2011. Genetics of educational attainment and the persistence of privilege at the turn of the 21st century [unpublished manuscript].
Pam, A., Kemker, S., Ross, C., Golder, R., 1996. The equal environments assumption in mz-dz twin comparisons: an untenable premise of psychiatric genetics? Acta Geneticae Medicae Et Gemellologiae 45, 349–360.
Penninkilampi-Kerola, V., Moilanen, I., 2005. Co-twin dependence, social interactions, and academic achievement: a population-based study. Journal of Social and Personal Relationships 22, 519–541.
Plomin, R., Willerman, L., Loehlin, J.C., 1976. Resemblance in appearance and the equal environments assumption in twin studies of personality traits. Behavior Genetics 6, 43–52.
Richardson, K., Norgate, S., 2005. The equal environments assumption of classical twin studies may not hold. British Journal of Educational Psychology 75, 339–350.
Rodgers, J.L., Kohler, H.P., Kyvik, K.O., Christensen, K., 2001. Behavior genetic modeling of human fertility: findings from a contemporary danish twin study. Demography 38, 29–42.
Rose, R.J., Koskenvuo, M., Kaprio, J., Seppo, S., Langinvainio, H., 1988. Shared genes, shared experiences, and similarity of personality: data from 14,288 adult Finnish co-twins. Journal of Personality and Social Psychology 54, 161–171.
Rowe, D., 1995. The Limits of Family Influence: Genes Experience and Behavior, 1st ed. Guilford Press.
Ryff, C., Almeida, D.M., Ayanian, J.S., Carr, D.S., Cleary, P.D., Coe, C., Davidson, R., Krueger, R.F., Lachman, M.E., Marks, N.F., Mroczek, D.K., Seeman, T., Seltzer, M.M., Singer, B.H., Sloan, R.P., Tun, P.A., Weinstein, M., Williams, D., 2011. National survey of midlife development in the united states (MIDUS II), 2004–2006.
Scarr, S., 1968. Environmental bias in twin studies. Eugenics Quarterly 15, 34–50.
Scarr, S., Carter-Saltzman, L., 1979. Twin method: defense of a critical assumption. Behavior Genetics 9, 527–542.
Schnittker, J., 2008. Happiness and success: genes, families, and the psychological effects of socioeconomic position and social support. American Journal of Sociology 114, 233–259.
Schnittker, J., 2010. Gene–environment correlations in the stress–depression relationship. Journal of Health and Social Behavior 51, 229–243.
Silventoinen, K., Sammalisto, S., Perola, M., Boomsma, D.I., Cornes, B.K., Davis, C., Dunkel, L., De Lange, M., Harris, J.R., Hjelmborg, J.V., 2003. Heritability of adult body height: a comparative study of twin cohorts in eight countries. Twin Research 6, 399–408.
Smith, K.B., Hatemi, P.K., 2012. OLS is AOK for ACE: a regression-based approach to synthesizing political science and behavioral genetics models. Political Behavior 35, 383–408.
Smith, R.T., 1965. A comparison of socioenvironmental factors in monozygotic and dizygotic twins testing an assumption. In: Vandenberg, S.G. (Ed.), Methods and Goals in Human Behavior Genetics. Academic Press, New York.
Towne, B., Czerwinski, S.A., Demerath, E.W., Blangero, J., Roche, A.F., Siervogel, R.M., 2005. Heritability of age at menarche in girls from the fels longitudinal study. American Journal of Physical Anthropology 128, 210–219.
Truett, K.R., Eaves, L.J., Meyer, J.M., Heath, A.C., Martin, N.G., 1992. Religion and education as mediators of attitudes: a multivariate analysis. Behavior Genetics 22, 43–62.
Turkheimer, E., 2000. Three laws of behavior genetics and what they mean. Current Directions in Psychological Science 9, 160.
Vinkhuyzen, A.A.E., Pedersen, N.L., Yang, J., Lee, S.H., Magnusson, P.K.E., Iacono, W.G., McGue, M., Madden, P.A.F., Heath, A.C., Luciano, M., Payton, A., Horan, M., Ollier, W., Pendleton, N., Deary, I.J., Montgomery, G.W., Martin, N.G., Visscher, P.M., Wray, N.R., 2012. Common SNPs explain some of the variation in the personality dimensions of neuroticism and extraversion. Translational Psychiatry 2, e102.
Wilson, P., 1934. A study of twins with special reference to heredity as a factor determining differences in environment. Human Biology 61, 629.
Xian, H., Scherrer, J.F., Eisen, S.A., True, W.R., Heath, A.C., Goldberg, J., Lyons, M.J., Tsuang, M.T., 2000. Self-reported zygosity and the equal-environments assumption for psychiatric disorders in the vietnam era twin registry. Behavior Genetics 30, 303–-310.