The Pennsylvania State University

The Graduate School

College of the Liberal Arts

WHAT CAN WE LEARN FROM TWIN STUDIES?

A COMPREHENSIVE EVALUATION OF THE EQUAL ENVIRONMENTS ASSUMPTION

A Dissertation in

Sociology

by

Jacob Felson

Submitted in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

December 2009

The dissertation of Jacob Felson was reviewed and approved* by the following:

John D. McCarthy
Professor of Sociology
Head of the Department of Sociology
Dissertation Advisor

Alan Booth
Professor of Sociology

Roger Finke
Professor of Sociology and Religious Studies

Katerina Bodovski
Assistant Professor of Education

*Signatures are on file in the Graduate School.

Abstract

Most evidence about the effects of genes on behavior comes from twin studies. The classic twin study method compares the similarity of monozygotic (MZ) twins on some trait with the similarity of dizygotic (DZ) twins on that same trait. If the rearing environments of MZ and DZ twins are analogous, then the magnitude of genetic influence on variation in a trait can be estimated based on the extent to which the correlation for MZ twins exceeds the correlation for DZ twins. If the rearing environments of MZ and DZ twins are not analogous, then the estimates of heritability generated from the classic twin study method may be confounded with differences between MZ and DZ twins in the similarity of their rearing environment. It has long been established that MZ twins grow up in a more similar environment than do DZ twins. However, behavior geneticists have generally asserted that the greater degree of similarity in the rearing environments of MZ relative to DZ twins does not bias the results of twin studies. Behavior geneticists claim that the similarity in outcomes is largely unaffected by similarity in rearing environments for MZ and DZ twins. This claim is known as the equal environments assumption (EEA).

In this paper, I argue that the empirical support for the validity of the EEA is not as strong as some behavior geneticists have claimed. I reanalyze some of the data on which ritually cited research on EEA is based, and I find that the evidence is more mixed than latter-day interpretations have suggested. I expose some of the flaws in the research that is frequently cited by researchers in support of claims about genetic influence on behavior.

Finally, I conduct the most comprehensive evaluation of the equal environments assumption to date. My analysis incorporates a larger and more diverse set of outcome variables than any previous research on the EEA. In my analysis, I also use a more extensive set of controls for environmental similarity than any previous study. Finally, unlike most if not all previous research on EEA, I generate estimates of genetic influence with and without controls for environmental similarity, in order to show the extent of EEA violation in a straightforward manner.

The findings are not easily categorized as supporting or undermining the EEA. Some violations of EEA were found for some outcomes, but the bias caused by these violations is likely to be modest in most cases. Researchers should not conduct twin studies without including controls for environmental similarity. However, the findings also indicate that many of the results of twin studies are robust with these controls. The flaws of twin studies are not fatal, but rather seem no worse (and maybe better) than the flaws of the typical causal study that relies on observational data.

Contents

List of Tables

List of Figures

Acknowledgements

I want to thank Professor John McCarthy, who was extremely effective as chair of my dissertation committee. John's guidance was instrumental in helping me finish this project. John was also very effective in helping me organize my ideas. My thesis benefited greatly from his editorial comments.

1 Introduction

Most of what we think we know about the power of genes on human behavior comes from studies of twins. In the standard twin study, researchers compare the level of similarity between identical twins to the level of similarity between fraternal twins on some characteristic of interest. Since identical twins are genetic clones, and fraternal twins share about 50% of their genes, the higher levels of similarity in psychology and behavior typically observed between identical twins, relative to fraternal twins, can be interpreted as evidence for genetic effects on psychological and behavioral traits. Since there is still relatively little direct evidence linking specific genes to specific behaviors, twin studies remain the main source of information about genetic influences on behavior.

Nonetheless, the validity of twin studies rests on an unproven assumption. The issue is that, as twin researchers have long acknowledged, identical twins receive more similar treatment than fraternal twins. Moreover, identical twins spend more time together than fraternal twins, especially in the formative early years. These facts do not necessarily pose a problem for twin studies, but may seriously undercut the results of twin studies if the higher levels of psychological and behavioral similarity observed between identical twins is partly the result of similarity in socialization. All of the results of classic twin studies, and hence much of what we think we know about the power of genes, rests on the assumption that higher levels of similarity in socialization between identical twins do not cause higher levels of similarity in outcomes. This assumption is known as the equal environments assumption (EEA), and this is the idea I evaluate in this paper.

There has been a considerable amount of research evaluating the equal environments assumption for different kinds of psychological traits and behaviors. Most twin researchers argue that there is overwhelming support for the EEA. But the findings about the EEA have not always been clear-cut, and it is difficult to make general conclusions about EEA without a thorough review and critique of the literature. No one has undertaken a comprehensive evaluation of the EEA using a nationally representative sample of twins.

In this project, I attempt to fill these gaps in the research. I do this in two ways:

1. I synthesize and critically evaluate research testing the equal environments assumption. In some cases, I show that the literature offers less support for the validity of twin studies than is often reported. Where possible, I reanalyze data used by previous researchers who evaluated the EEA.

2. I test the EEA in a comprehensive way using a nationally representative sample

of twins. I examine whether the apparent effects of genes on a large variety of characteristics are confounded with environmental similarity. The study incorporates a broader array of measures of environmental similarity, and a greater number of outcomes, than nearly any previous study.[1]

My synthesis of work testing the EEA shows the traits where it is robust, and the traits where EEA may be problematic. At present, researchers predisposed for and against the behavior genetic paradigm tend to argue that the EEA is either valid or invalid. As is often the case with research findings, the distribution of evidence is more evenly distributed than the interpretations of the evidence. My paper concludes that it is flawed to view the EEA as a singular hypothesis to be accepted or rejected. Some twin researchers have argued that they do not need to control for environmental similarity because previous studies have found evidence supporting the EEA (e.g. Alford et al., 2005). My review of the literature suggests that this assertion is problematic. It is justifiable to omit a control variable X in a study of the causes of some variable A simply because X was not a significant predictor in a previous study of the causes of another variable, B. Even when variables A and B are similar, or A and B measure the same concept, social researchers are likely to control for X if there are theoretical reasons to believe it could confound the effects of independent variables.

The bottom line is that there should be no EEA: equal environments should be tested for rather than assumed. Then researchers can move beyond more abstract debates about the theoretical viability of the EEA, and focus more on developing methodologically rigorous ways of testing it.

1.1 Plan for the present study

The present study is organized in the following fashion. In sections 1.2 and 1.2, I explain the twin study methodology, and talk about how researchers have used it. In section 1.3, I evaluate how well the classic twin study method isolates genetic causes.

Section 2 reviews the literature that evaluates the equal environments assumption. In some cases, I reanalyze the data from these studies. Section 2.1, I review the earliest studies on the EEA that established that MZ twins experience more similar environments than do DZ twins. Section 2.2 evaluates early studies that examined whether genetic effects found in twin studies are confounded by environmental effects. In section 2.2, I briefly review recent studies that evaluated the equal environments

[1] Only one study offers as comprehensive an analysis of EEA as the present study (Loehlin and Nichols, 1976), but this study was not conducted on a nationally representative sample of twins.

assumption. In the final part of the Literature Review, 2.4, I discuss the bearing of other types of genetically informed research on the classic twin study methodology.

Section 3 begins the original analysis of the present study. The data and methods of the original analysis are described in sections 3 and 4, respectively. The results of the original analysis are presented in section 5. In section 6, I review the implications of the critical review and reanalyses of previous research, and of the results of the original analysis.

1.2 Estimating the effects of genes on variation in traits with twin studies

Our present thinking about genetic causes of behavior is based on indirect evidence. Scientists have not isolated the mechanism by which fat causes heart disease, and neither have scientists isolated the mechanisms by which genes cause any of the complex behaviors that they are purported to cause: criminal tendencies, high performing brains, or the expression of personality characteristics of various kinds.

Instead, much of what scientists believe about genetic causation stems from studies comparing monozygotic (MZ) twins with same-sex dizygotic (DZ) twins. The comparison between MZ and DZ twins would seem to be an ideal way of separating out the effect of genes from the effect of environments. Monozygotic twins emerged from one egg, so they are genetic clones of each other. Dizygotic twins, on the other hand, emerged from two eggs, so they are no more genetically similar to each than are non-twin siblings. Yet unlike non-twin siblings, both MZ twins and DZ twins develop in parallel with each other.

So we may be able to assume that MZ twins and DZ twins share similar, if not equivalent, environments, at least insofar as both pairs of twins grow up together. In that case, the critical difference between MZ and DZ twins is genetic: MZ twins share 100% of their genes, while DZ twins share (on average) only 50% of their genes.

Estimating heritability

Heritability is defined as the proportion of variation in a trait that is due to variation in genes. Heritability is often abbreviated as h^2. An estimate of heritability can be obtained quickly by doubling the difference between the correlations on a trait for MZ and DZ twins. For example, a simple heritability estimate for a trait that correlates at 0.5 between MZ twins and 0.25 between DZ twins would be $2 \times (0.5 - 0.25) = 0.50$.

The reason heritability can be estimated in this way can be seen from the decomposition of the MZ and DZ correlations displayed below, where r_{MZ} is the correlation for MZ twins, r_{DZ} is the correlation for DZ twins, h^2 is heritability (as noted above) and c^2 is the effect of the shared environment.

$$r_{MZ} = h^2 + c^2 \tag{1}$$
$$r_{DZ} = .5h^2 + c^2 \tag{2}$$

Because MZ twins reared together share 100% of their genes and 100% of their family environment, the correlation between MZ twins on any trait is equivalent to the sum of parameters representing genes and environment. DZ twins share a family environment in the same way, but share, on average (disregarding assortative mating) 50% of their genes. So the correlation for DZ twins is decomposed accordingly. By subtracting the second equation from the first, we can see that the difference between the correlations is equal to $.5h^2$. Twice the difference is thus an estimate of h^2.

Following this logic, scientists are able to attribute greater correlations in behavior between MZ twins relative to DZ twins to genetic differences. Behavior geneticists have found that MZ twins are indeed more similar to each other than are DZ twins to each other on a great number of behaviors. For example, the IQs of MZ twins correlate to a much greater extent than the IQs of DZ twins (Jensen, 1998, pp.445-458). The same can be said to a lesser extent for personality characteristics that interest psychologists, such as openness to experience, conscientiousness, extraversion, agreeableness, and neuroticism (Rowe, 1994, p.65).

Differences in concordance rates between MZ and DZ twins have also been found for traits of interest to sociologists. Using the twin method, political scientists have recently claimed to find evidence that political attitudes – ranging from very liberal to very conservative – also have genetic origins (Alford et al., 2005). Economists have found genetic effects on educational achievement, and criminologists have found genetic effects on criminal behavior. The list goes on ad nauseam. For almost every behavior that could plausibly be influenced by genes – even in a very indirect way – scholars have claimed to have found such effects using the twin method.

From behavior geneticists' point of view, the next step is to isolate the particular genes responsible for influencing behavior. Given the evidence from twin studies, many behavior geneticists and scholars sympathetic to their point of view believe these behavior-influencing genes will eventually be found. Alternatively however, we might pay closer attention to some of the assumptions that undergird the twin studies, in order to see whether such studies can be trusted to provide answers for which direct validation is forthcoming.

1.3 Evaluating the twin study as a natural experiment

Before examining the assumptions of twin studies specifically, I want to consider the twin study as an experimental method. When searching for causes, scientists generally consider experiments to be the gold standard. Observational studies, meanwhile, are a distant second best, and only dominate social science because ethical and practical issues preclude most kinds of social experimentation. Thus, to the casual observer, twin methods seem superior to regression. By this standard, we might assume that behavior geneticists are working from a stronger foundation than are most social scientists.

Let us evaluate the twin study method in terms of how well it isolates causes. A good way to do this is to consider the counterfactual model of causality. The counterfactual model of causality is a heuristic for thinking about causation championed by prominent statisticians such as Gary King (King et al., 1994). When thinking about counterfactuals, we think about what would have happened to the dependent variable if the causal variable had taken on a different value, yet everything else had remained the same. We would need to compare the outcomes for a subject with the outcomes to a clone of that subject who is modified in only one way. For example, if we are interested in whether a particular gene causes higher IQ, we would have to imagine a scenario in which we could travel back in time and manipulate the gene in question. If we could accomplish both of those tasks, we could observe whether the gene affected IQ.

According to the counterfactual model of causality, every effort to find causes should attempt to simulate the counterfactual scenario. Moreover, every method that attempts to isolate causes should be evaluated in terms of how well it simulates the counterfactual. The two most common social science methodologies – experiments, and multiple regression using observational data – simulate the counterfactual in different ways.

Experiments simulate the counterfactual through application of the law of large numbers to groups of subjects. If the scientist includes a substantial number of people in her study, and randomly assigns subjects to experimental and to control conditions, then the experimental group should be identical to the control group in terms of all relevant characteristics. In accordance with the law of large numbers, differences between two groups of people chosen at random will decrease the number of subjects increases. If there are no differences between treatment and control groups, the groups are essentially clones of one another. Thus, a comparison of the treatment group with the control group should yield similar information to a comparison of a person who received a treatment with that same person who did not receive the treatment. Through the alchemy of probability theory, the ideal experiment converts leaden correlations into golden causes.

The other common social science methodology, multiple regression using observational data, is typically viewed as an inferior alternative to experiments. Regression allows researchers to simulate the counterfactual by comparing people who are similar on all of the ways that researchers can measure them. With regression, researchers essentially compare individuals who provided the same answers on the survey. But of course surveys are not make-it-from-scratch recipes for human beings. Even if a survey asked every question imaginable, the existence of survey response error makes such a thing elusive.

Instead, the extent to which regression simulates the counterfactual will be dependent on the researcher's knowledge of relevant variables, as well as the quality of survey measurement. There is ambiguity inherent in the regression methodology. Unlike the well-run experiment, the chance of differences between treatment and control groups does not become negligible as the number of people in the study increase.

Twin studies have typically been treated as experiments by behavior geneticists. Perhaps for this reason, researchers have been more willing to give twin studies the benefit of the doubt than regression results. Doubters of regression results can almost always come up with some relevant yet excluded variable that, if controlled, would obliterate the main finding. Yet the well-run experiment is not so easily dismissed. Twin studies enjoy this privilege.

When researchers have published regression studies showing something that seemed preposterous, many scholars were able to call the results into question by citing a few excluded yet relevant controls. For example, when suicide researcher Stack (1992) published a study based on aggregate data claiming that listening to country music heightened the risk of suicide, skeptics were quick to point out flaws in the aggregate-level regression. Yet the results of twin studies indicating that parents do not influence their children's personalities independent of genetic inheritance do not seem to have prompted great methodological critique or revision within the field of behavioral genetics.

The more counterintuitive a result, the more energy should be spent evaluating the underlying methodology. This seems to be the case with regression results, in part due to the underlying ambiguity in that methodology. But as I have said, the same does not appear to be true with twin studies. Instead, the counterintuitive results of many twin studies have been followed by more twin studies, rather than more evaluation of twin studies' underlying claim to causality. Following several decades of twins studies showing very modest influence of family environment on personality, Harris (1999), in The Nurture Assumption, concluded that parents' behavior does not influence their children much. According to Harris, parents have little influence outside their genes, provided a certain minimum threshold of care, and a lack of abuse.

Let us examine the extent to which twin studies simulate the counterfactual. Consider the controversial case of IQ scores. If one is interested in the effects of genes on IQ scores, then one imagines two individuals who are identical in every way except that one has genetic material hypothesized to produce high IQ (call him the "treated") and the other had genetic material hypothesized to produce low IQ (call him the control). If we find that the treated individual scores higher on IQ tests than the control individual, we can conclude with certainty that our hypothesis is correct.

Even in this counterfactual situation, the precise mechanism of influence may not be clear. We would not be able to assume that the gene affected speed of brain processing, the biochemistry directly influencing performance on an IQ test. Instead, the gene could affect someone's interest in abstract thinking, or one's level of motivation and self-control that permit one to develop the skills necessary to perform well on an IQ test. It is worth keeping in mind that even the counterfactual would not give us exactly what we wanted.

Even though the twin study counterfactual is not exactly what we want, it is still a good goal for which to strive. How does the twin study stack up on this account? To simplify matters, I consider only two pairs of twins, one monozygotic and the other dizygotic. In the language of experiments, the monozygotic pair is the treatment group, while the dizygotic pair is the control group. The monozygotic pair of twins should be identical to the dizygotic pair of twins in every respect except that the monozygotic twins are genetic clones of each other, and the dizygotic twins are related like non-twin siblings.

Twin studies seem to simulate the counterfactual rather well. Almost every difference between MZ and DZ pairs that one can imagine ultimately stems from their genetic differences. The greater similarity observed in MZ twins than observed in DZ twins may result from the fact that parents and peers treat MZ twins more similarly than they treat DZ twins. But the similarity of treatment received by MZ twins can ultimately be traced back to their genetic relatedness.

To take a concrete example, consider how twins' IQ may be indirectly affected by differences in attractiveness between twins, which will of course generally be much smaller for MZ twins than for DZ twins. The genetic effect on appearance is presumably direct, and relatively uncontroversial.

So how might twin similarities in appearance affect twin similarities in IQ? Research has shown that attractive people often receive better treatment than less attractive people. It has also been shown that more attractive people are widely perceived to be more intelligent than less attractive people (Dion et al., 1972). In accordance with these findings, parents may treat their more attractive children better than their less attractive children. Parents may also unconsciously assume that their more attractive child is more intelligent than their less attractive child, especially early in their child's life, before more direct evidence of intelligence is apparent. (In general, perceptions of intelligence rely more heavily on attractiveness when there is little relevant information on which to make an evaluation.) Differences in parental treatment may be subtle in many cases, manifested as a lower level of parental encouragement, rather than as actual maltreatment. Differences in parental treatment may affect children's motivations to pursue difficult problems, such as the sorts of problems that appear on IQ tests.

This is only one of the many ways in which MZ genetic equivalence may lead to similarity in IQ. The pathways are numerous. Nonetheless, as I have said, the twin method appears to simulate the counterfactual quite well. All the ways in which MZ twins are treated more similarly than DZ twins, as well as all of the ways in which MZ twins' interactions differ from DZ twins, stem from the genetic equivalence of the MZ twins and the genetic non-equivalence of the DZ twins.

So the problem with twin studies is not their purchase on causality per se. Unlike the typical experiment, twin researchers need not be concerned about pre-existing differences between treatment and control groups. In the twin 'natural experiment,' genes are the treatment, and a person's genes remain constant from the moment of conception.

Instead, twin studies suffer from some of the same problems that plague laboratory experiments. Like many laboratory experiments, twin studies simulate a particular counterfactual very well, but the counterfactual they are simulating is not the one we are most interested in. In other words, twin studies may suffer from external validity problems.

More specifically, the problem is that the similarity in outcomes between MZ twins may be maximized in many ways that have specifically to do with the nature of twinship. If the heightened similarity of MZ twins relative DZ twins is partly explained by the special experience of growing up with a clone of oneself, then the results of twin studies may not apply to non-twin populations. Since twins themselves only account for about 2% of births, the failure of twin studies to generalize to the non-twin population means that that twin studies may be very limited in their usefulness. (Although the percentage of twins is increasing due to the rise in the use of fertility drugs by women in their forties, it is highly unlikely that twin births will encompass more than a negligible percentage of overall births any time soon.)

In fact, if researchers were convinced of the problems posed by the unique MZ twin environment, they may have long since abandoned, or at the very least modified, their twin study method.

Put another way, the problem here is that genetic influences may be a necessary but insufficient cause of differences in concordances between MZ and DZ twins. To the extent that the idiosyncrasies of monozygotic twinship explain the higher MZ correlations, twin studies lack external validity.

Earlier I discussed a central assumption underlying twin studies, an idea known as the equal environments assumption (EEA). According to the EEA, the environments of MZ twins are not more similar than the environments of DZ twins in any way that causes MZ twins to turn out more similar to each other than DZ twins turn out.

Violations of EEA compromise the external validity of twin studies. If the EEA is compromised, then we cannot assume that the genetic effects revealed in twin studies are present in the non-twin population. Since the foundation of the twin study methodology is built on the equal environments assumption, this study is devoted to the task of evaluating whether EEA is sound.

Before presenting my own analyses, I review previous evaluations of the EEA.

2 Literature Review

The literature review is broken into four sections. First, in subsection 2.1, I review the earliest studies conducted between 1930 and 1960, which established that MZ twins experience more similar environments than DZ twins. Next, in subsection 2.2, I provide in-depth review and in some cases reanalysis of studies published between 1965 and 1979. In that section, I critically evaluate the studies that have been cited most frequently in support of the idea that twins' environmental similarity does not correlate with similarity of outcomes. The next subsection, 2.3, I synthesize findings of studies of the EEA published in the last thirty years. Since I discuss most of the key issues with studies of the EEA in subsection 2.2, my review of more recent studies is brief. Finally, in the last section of the literature review, 2.4, I briefly discuss the evidence from studies of twins reared apart and from studies of adopted children that corroborate the results of classic twin studies.

2.1 Earliest studies (1930–1960) establish that MZ twins experience more similar environments than DZ twins

Criticism of the equal environments assumption is almost as old as the classic twin method itself. The method of parsing out the influence of nature versus nurture was introduced by Hermann Siemens in a 1924 book, Die Zwillingspathologie (Joseph, 2004, p.19).[2] As far as I can establish, American eugenicist Holmes (1930) was the first to consider the possibility similarity of experience could explain similarity of outcomes of MZ relative to DZ twins. The same problem was recognized by a medical statistician, Stocks (1930).

Problems with EEA continued to be raised by others. I review and reanalyze the most extensive early study below.

2.1.1 Review and reanalysis of Wilson (1934)

While problems with the EEA had been raised in print only six years after the twin study was introduced, the first comprehensive quantitative test of the EEA was conducted by Wilson (1934). Wilson (1934) seems to have been the first to evaluate the equal environments assumption in a systematic way with data. Because this study is the first of its kind, I include an extensive discussion and re-analysis of the results here.

The study was conducted by the Institute of Child Welfare at the University of California. The investigators attempted to gather a representative sample of adolescent twins by canvassing nineteen junior and sixteen senior high schools from four cities: Alameda, Berkeley, Oakland and San Francisco. Twins in these schools were identified through school administrators. Investigators tried to ensure that twins who attended different classes or different schools were included.[3]

Wilson (1934) interprets greater similarity between MZ twins than DZ twins on a

[2]Many have credited Francis Galton with inventing the twin method. While Galton was the first to study twins as a window into heredity, he did not compare MZ and DZ twins. Galton examined similarities between twins, and between family members, but he could not separate the effects of nurture from nature, as zygosity was not yet discovered. In fact, Galton died in 1911, at least a decade before the distinction between MZ and DZ twins became well-established (Joseph, 2004, p.17)

[3]It is important to include twins who attend different classes or different schools, given that the focus of the study is on twins' environmental similarity. As far as I could discern from the article, the investigators failed to find the co-twin in only 12 (about 6%) of the 206 twin pairs identified. According to Wilson, most of these pairs were different sex fraternal twins, who will not be examined here. Thus, it is unlikely that the results I present from this study are affected much at all by selection bias.

variety of personality characteristics as evidence that environment differs by zygosity. In this way, Wilson (1934)'s work contrasts sharply with later twin studies. In an earlier article, Jones and Wilson (1932) argued that the hypothesis of "equivalent average differences in the post-natal nurture of the two types of twins [was] unsound." Based on the fact that MZ twins were more similar than DZ twins, the authors argued that the normative expectations for MZ twins to behave in a very similar way relative to DZ twins will have an impact on the behavior of the twins. Yet later twin researchers interpret the *same phenomenon* – the relationship between zygosity and various characteristics in adolescence – as evidence for the effect of genes. The same type of evidence that would later be taken as evidence of the utility of twin studies to identify genetic effects is interpreted in early 1930s as a challenge to the foundations of twin studies.

When examining a study as old as Wilson (1934), measurement of zygosity is a potential concern. After all, the study was completed before the discovery of DNA (Watson and Crick, 1953), so no authoritative tests of zygosity were available for use either as direct measures or as a means of validating questionnaire methods. Moreover, the study was conducted before techniques for discerning zygosity based on self-reports of physical similarity had been developed and validated.[4] Nonetheless, the method by which authors of the study determined zygosity appears to be both highly reliable and highly valid.

In order to determine zygosity, investigators relied mostly on anthropometric measures. In order for twin pairs to be classified as identical, the pair had to have the same eye color, hair color, ridges in the fingers, and be judged to be strikingly similar in overall appearance. There were other criteria as well. Reliability, as measured by the level of agreement between independent raters, appears to have been high. Another scholar's independent classification of a subset of eighty pairs in the sample was discrepant with the classification used by Wilson (1934) in only three out of eighty pairs examined. These discrepant cases were classified as undetermined and left out of the analysis.

Wilson (1934) compared the degree of environmental similarity reported by MZ twins with the degree of environmental similarity reported by DZ twins. Comparisons were made on MZ twin pairs and DZ twin pairs matched on age, gender, and socioeconomic status. Twins were asked a variety of questions about the extent to which they were raised in similar environment as their co-twin. Here, I reanalyze all questions that asked about environmental similarity of twins. Due to the age of the study, I do not have access to the original data; thus, my analyses are based solely

[4] Techniques that discern zygosity based on self-reports of physical similarity have been shown to be over 90% accurate (Kasriel and Eaves, 1976).

on the data presented in Wilson (1934). The questions that I chose are presented in table 1 below. Answer categories offered to respondents are presented in the column on the right.

Measures of environmental similarity differed in terms of the number of response categories. Fourteen of the questions had three or four categories that could be ordered from least similar to most similar environment. Two other questions, about waking time and bedtime, included two responses that were equivalent in terms of similarity of environment. Categories that were equivalent in terms of environmental similarity were collapsed. There was one additional question about whether the twins attended the same school that was dichotomous.

Both twins were asked these questions independently via paper questionnaire, although apparently twins were instructed to confer with their co-twin to verify their answers. Unfortunately, Wilson (1934) was not clear about how he handled cases in which there was disagreement between twins. One is led to presume that twins agreed in their responses in most cases, given that they were instructed to confer with each other. In any case, Wilson (1934) provides no analysis of intra-pair reliability.

A study of this kind would be conducted very differently today. For example, researchers would probably not allow twins to confer with each other about their answers to survey questions. While the choices made by Wilson (1934) were not ideal by contemporary standards, there are merits in the choices that were made. The fact that the twins were encouraged to discuss their answers may increase the validity of the survey responses. For example, twins may refresh each other's memory.

The relevant question here is whether the process of conferral affects MZ twins differently than it affects DZ twins. Encouragement of twins to compare answers could increase MZ similarity relative to DZ similarity. The process of conferral could increase the power of social norms on twins' responses. MZ twins who may disagree about the similarity of their environment when answering questions in isolation from each other, may, when allowed to reconcile their responses, be more influenced by expectations about MZ sameness, and indicate higher levels of similarity. Norms of MZ similarity may be more active when twins are conferring about their answers than when they are answering questions in isolation.

DZ twins may not reconcile their responses in the direction of greater closeness to the same extent as MZ twins. The effects of conferral on survey questions may not be as strong for DZ twins as for MZ twins because social expectations about DZ similarity are likely to be weaker than social expectations about MZ similarity. Hence, DZ twins may not be as likely as MZ twins to reconcile their answers in the direction of increased closeness. While this pattern is a matter of speculation and

Table 1: Descriptive statistics for reanalysis of Wilson (1934)

Topic	**Question**	**Answers**
School	Twins indicated whether they attended the same or different schools.	same, different
Bedtime	Does your twin usually go to bed (before, at the same time, or after) you do?	same, different
General	What is the longest time you were ever separated from your twin?	≤1 day, 2-8 days, 8 or more days
School-to-home	Do you and your twin go home from school together?	always, sometimes, seldom
Waking time	Does your twin usually get up (before, at the same time, or after) you do?	same, different
General 2	How many times have you been separated from your twin for more than a day?	0, ≥3x, 4x
Group	Do you and your twin go around with the same group?	always, sometimes, no
Clothing preferences	If you dress like your twin, do you like to be dressed like your twin?	yes, indifferent, no
Errands	If you are sent on an errand does your twin go with you?	always, sometimes, seldom
Clothing	Do you dress like your twin?	always, usually, seldom
Friends	Do you have the same chums?	yes, sometimes, no
Lessons	Do you study your lessons with your twin?	usually, sometimes, seldom
Sleeping	Do you and your twin sleep in different bedrooms, different beds in the same room, or in the same bed?	different bedrooms, different beds, same room, same bed
Classes	What classes do you have together with your twin?	all, almost all, part, none
Prefer twin	Would you rather be with your twin than with anyone else?	yes, sometimes, no
Compared to sibs	Do you think you are together with your twin more than ordinary brothers and sisters are together?	yes, about the same, no
Mother preference	Does your mother like to have you be with your twin most of the time?	yes, doesn't care, no
Movies	If you go to a movie does your twin go with you?	always, sometimes, seldom

cannot be tested, it is useful to keep in mind as one interprets the results of my reanalysis of Wilson (1934). Since twins did not answer the questions on their own, the study may exaggerate the differences in the degree of environmental similarity experienced by MZ twins relative to DZ twins.

The original analysis consisted of a series of contingency tables. I present the data in a more parsimonious way. In order to compare the effects of zygosity on the similarity of different measures of environment, I ran loglinear models separately for each measure of environmental similarity. I employ a type of loglinear model called the column effects model (Powers and Xie, 2000, p.122) that is suited for estimating the relationship between a nominal variable and an ordinal variable.[5] In the standard loglinear model, there is a parameter for each (non-redundant) cross-classification of variable categories. By contrast, in a column-effects model, the relationship between a dichotomous variable (zygosity) and an ordinal measure (of environmental similarity) is represented by a single parameter. The column-effects model is hence considerably more parsimonious than the standard log linear model. Also, the column-effects model facilitates comparisons between the magnitudes of the effects of zygosity on ordinal measures of environmental similarity that have unequal numbers of categories.[6]

The parsimony of the column-effects model comes at the cost of an assumption that cannot be tested in the present analysis. In the column-effects models I estimate here, the gaps between categories of environmental similarity are assumed to be equivalent. This assumption is necessary, since there are not enough degrees of freedom in these models to test models in which the assumption is relaxed. However, the conclusions I reach here are unlikely to be sensitive to violations of this assumption.

The data on which these models are based are cross-classifications (threeway crosstabs) of each environmental measure by zygosity and by gender. These cross-classifications were constructed based on the percentages and sample sizes listed in Wilson (1934). Given that the sample size was relatively small, some cells contained

[5] The only distinction between column-effects and row-effects models is the placement of the ordinal variable and the nominal variable in the contingency table. Since the choice of which variable to place in the rows and which variable to place in the columns is arbitrary, the distinction between "column-effects" model and "row effects" models in the literature is not particularly useful. The models are fundamentally the same.

[6] The alternative would be to represent the relationship between zygosity and each environmental measure by a set of odds ratios. For example, the relationship between zygosity (MZ vs. DZ) and a three-category environmental measure could be expressed with two (non-redundant) odds ratios. Likewise, the relationship between zygosity and a four-category environmental measure could be expressed with three odds ratios.

no observations. Zero cells are problematic in analysis of ratios since ratios with denominators of zero are undefined. In the present analysis, 8 out of 208 cells in the data had no observations. I replaced zero cells with a small constant of 0.5, as recommended by Knoke and Burke (1980, p.64). According to Knoke and Burke (1980), this procedure, while ad hoc, is conservative and will tend to underestimate effects and their significance.

I estimated at least two models for each question listed in Table 1: (1) a gender-invariant model, in which the relationship between zygosity and environmental similarity is assumed to be constant across gender, and a (2) gender-specific model, in which the relationship between zygosity and environmental similarity is estimated separately for boys and girls. The equation for each of these models was:

$$\log n_{ijk} = \mu + x_i\beta_{xi} + \cdots + x_{J-1}\beta_{x(J-1)} + y_2\beta_y + z_2\beta_z + y_2z_2\beta_{yz} + y_2j\beta_{yj} \quad (3)$$

Equation 3 predicts the number of twin pairs in the cross-classification of each measure of environmental similarity × zygosity × gender. i indexes categories of the environmental variable, j indexes zygosity (MZ/DZ), and k indexes gender. Thus the outcome variable is the natural logarithm of the number of twin pairs in the i^{th} category of environmental similarity, the j^{th} zygosity, and the k^{th} gender. μ is the intercept and $x_i + \cdots + x_{J-1}$ are a set of dummy variables, one for each of $J-1$ categories of the measure of environmental similarity. y_2, and z_2 are dummy variables for dizygosity, and female, respectively. The coefficient for the multiplicative term between y_2 and z_2, β_{yz}, measures the relationship between zygosity and gender. This interaction term z_2z_2 is included since a greater percentage of MZ twins than DZ twins in the sample happened to be male.

The coefficient for the final term in the equation, β_{yj}, is the so-called "column-effect" that estimates the effect of zygosity on each measure of environmental similarity. When the environmental similarity variable is dichotomous, this term is simply the logarithm of the odds ratio reflecting the relationship between zygosity and environmental similarity.

Fit statistics for these models are presented in Table 2. The abbreviated question topics in the left column of Table 2 reference the question topics presented in Table 1.

The first two columns in table 2 display fit statistics and degrees of freedom for models in which it is assumed that the relationship between zygosity and environmental measures does not vary by gender. The next two columns contain fit statistics and degrees of freedom for models estimating gender-specific effects of zygosity on each environmental measure. The final column displays the results of likelihood-ratio

Table 2: Model fit statistics for reanalysis of Wilson (1934)

	Gender-invariant		Gender-specific		Likelihood ratio test	
Question	G^2	DF	G^2	DF	Diff in G^2	Sig.
School	3.148	3	7.25e-13	0	3.148	0.369
Bedtime	2.578	3	8.66e-15	0	2.578	0.461
School-to-home	0.129	3	.0542373	2	0.074	0.785
Longest separation	4.399	3	4.067772	2	0.331	0.565
Waking time	2.135	2	4.00e-15	0	2.135	0.344
Times separated	1.379	3	.2032331	2	1.176	0.278
Clothing preferences	2.051	3	1.916532	2	0.134	0.714
Group	9.453	5	9.158653	4	0.294	0.588
Errands	4.283	3	2.929097	2	1.354	0.245
Studying	2.595	3	2.437705	2	0.157	0.692
Sleeping	1.501	3	.2649668	2	1.236	0.266
Friends	4.411	3	2.798542	2	1.613	0.204
Clothing	6.329	3	6.043394	2	0.286	0.593
Classes	6.452	5	5.708449	4	0.744	0.389
Preference for twin	3.046	3	.3110396	2	2.735	0.098
Movies	1.393	3	.051789	2	1.341	0.247
Mother preference	2.245	3	.1115454	2	2.133	0.144
Compared to sibs	5.614	3	3.323387	2	2.290	0.130

tests comparing the model fit of the gender-invariant model with the model fit of the gender-specific model for each measure of environmental similarity.

According to the likelihood-ratio tests, the gender-invariant model fits the data for every measure of environmental similarity. The model fit of the gender-specific models is never significantly better than the model fit of the gender-invariant models. In the case of dichotomous environmental measures – school, bedtime, and waking time – the gender-specific model is the saturated model. In these cases, the gender-specific model has zero degrees of freedom and thus fits the data perfectly by definition.

If sample sizes were larger, we might find some significant interactions between zygosity and gender. For thirteen of the eighteen questions, the gender-specific model estimated a larger zygosity effect for females than for males. There is a trend here suggesting that the effect of zygosity on environmental similarity is greater among girls than among boys.

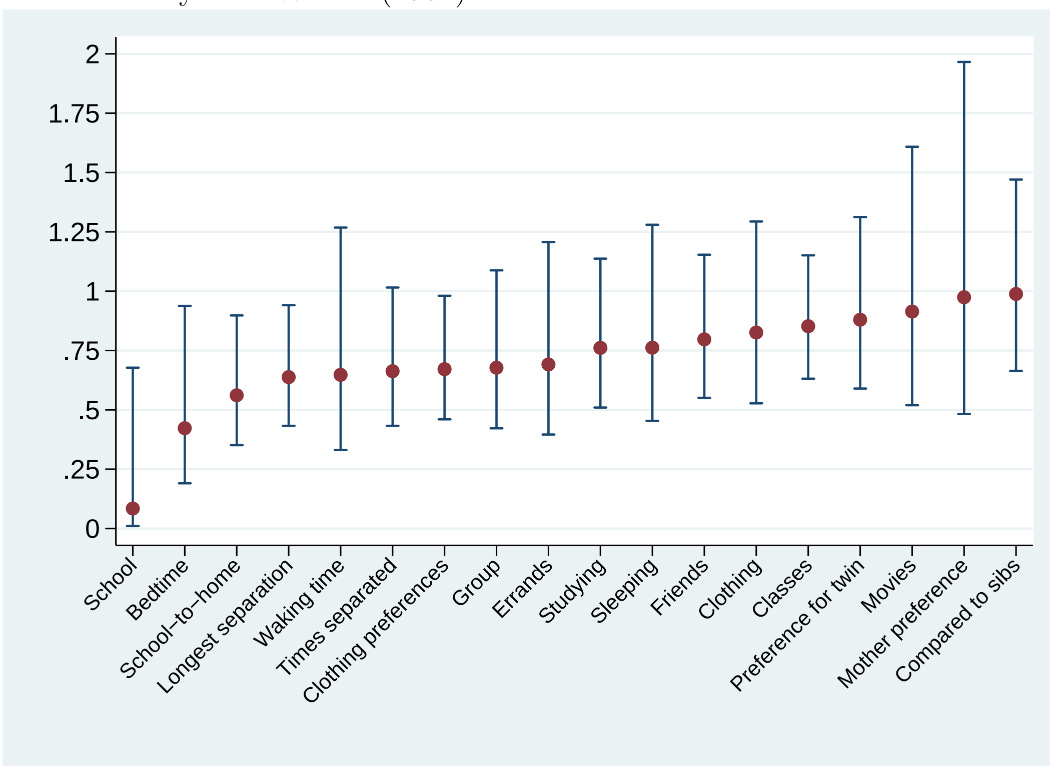
Since the interactions with gender were not significant, I do not report them here. Instead, I depict point estimates for the odds ratios of the column effects for each variable as well as associated 95% confidence intervals, in Figure 1. Measures of environmental similarity are arranged in descending order of magnitude of the zygosity effect.

For all but one question, DZ twins in the sample report their environment to be less similar than do MZ twins. To be sure, confidence intervals for the odds ratios include 1 for all but five questions. Zygosity has a significant effect on the following environments: bedtime, same school, clothing preferences, longest separation time, and transportation from school-to-home. Although significance is limited to five out of eighteen questions, the point estimates all trend in the same direction. It is likely that many more effects would have been significant had sample size been larger.

There is considerable overlap among the confidence intervals for different questions, so conclusions about the relative sizes of the effects are tentative at best. Nonetheless, the data are suggestive of several patterns. It seems that parents are particularly averse to separating MZ twins, relative to parents of DZ twins. When MZ twins are separated, they are separated for shorter time periods than DZ twins. In this context, separation could mean 'away camp', trips or boarding school.

The effect of zygosity on parents' tendency to keep twins together may simply be a reflection of genetic effects on twin behavior. For example, parents may send delinquent children away to reform school, and keep well-behaved children at home. Parents' propensity to keep MZ twins together may simply reflect parents' responses to MZ twins' genetically-derived behavioral similarity. On the other hand, relative genetic dissimilarity of DZ twin pairs relative to MZ twin pairs may induce rela-

Figure 1: Comparing environmental similarity of DZ twins to that of MZ twins, based on reanalysis of Wilson (1934)



tive dissimilarity in environments. Relative to MZ twins, DZ twins may be more likely to have formative experiences in isolation from their co-twin. Since the origin of environmental differences is uncertain, the validity of the equals environments assumption may also be uncertain. Only two out of the eighteen environmental questions elicited similar answers from MZ and DZ twins: Does your mother like to have you be with your twin most of the time? (abbreviation: Mother preference), and Do you think you are together with your twin more than ordinary brothers and sisters are together? (abbreviation: Compared to sibs). For both of these questions, respondents are asked to report (at least in part) about the experiences of others. Questions that require respondents to report about the attitudes and behavior of others are likely to be less reliable than questions that ask respondents to report on their own attitudes and behavior (Alwin, 2007, p.152). Differences on these variables hence may be weakened by their lower reliability. To be sure, the rearing environment of DZ twins may be much more similar to the rearing environment experienced by MZ twins than to the rearing environment of 'ordinary' siblings.

2.1.2 Early study comparing MZ and DZ twins by Mowrer (1954)

One of the earliest studies comparing the relationships between MZ twins to the relationships between DZ twins was conducted by a sociologist, Mowrer (1954). Mowrer analyzed a mail survey of 612 twins who were identified from records of high school graduates from 35 cities nationwide. Mowrer found some striking differences between MZ and DZ twins. When asked to identify the member of the family who would be missed most if he or she died, 49% of MZ twins chose their co-twin, while only 25% of same-sex DZ twins chose their co-twin. When asked to identify the member of their family whom they felt understood them most, 61% of MZ twins chose their co-twin, as compared with 39% of DZs. In addition, MZ twins identified more strongly with their co-twins than did DZ twins.

MZ relationships tended to be more positive than DZ relationships. Mowrer found that a minority (39%) of MZ twins reported that they were never jealous of their co-twin, while a majority (53%) of same-sex DZ twins reported an absence of jealousy for their co-twin. These findings, reported in a 1954 issue of the American Sociological Review, provide some of the first evidence that developmental dynamics may be different among (at least some) MZ twins than among DZ twins. Many MZ twins assume a role in the lives of their co-twins that is analogous to the maternal role in young children in terms of its level of intimacy and importance. Among DZ twins, this is less often the case.

Mowrer also provides evidence that MZ twins are treated much more similarly than are DZ twins. While the overwhelming majority (79%) of MZ twins reported receiving identical gifts from their parents, a minority (39%) of same-sex DZ twins reported receiving identical gifts from parents.

Mowrer's study provides early evidence that the MZ twin bond is sui generis. Nearly double as many MZs (27%) as DZs (14%) reported that they attracted considerably more attention from relatives and close friends of the family than did their non-twin siblings. The twinship bond – and thus the effects the bond has on behavior – may be strengthened when it is the focus of others' attention. And DZ twinship seems to be less often the focus of others' attention than MZ twinship.

To be sure, DZs were more similar to MZs in their perceptions of norms governing twin psychology. When asked whether they felt an obligation to be closer to their co-twin than to their other siblings, 70% of MZ twins and 56% of DZ twins answered affirmatively. In addition, the difference in wearing each other's clothes was small. 85% of MZ twins and 76% of DZ twins reported that they sometimes wore each other's clothes.

2.1.3 Other studies conducted prior to 1960

Further evidence that MZ twins share more similar environments than DZ twins was found by Smith (1965). Among adolescent twins, Smith (1965) found that MZ twins dressed alike more often, studied together more, and shared more close friends than did same-sex DZ pairs. Smith interviewed 164 twin pairs about work, school, sports, leisure, sleep, dress, study habits, and food and beverage preferences. Zygosity was related to similarity of environment, especially among girls. Female MZ twins may be more dependent on each other than male MZ twins, relative to DZ twins. MZ twins were more likely to have the same friends, and the same patterns of attendance at sports events. MZ twin girls were more likely to dress alike, but there was no relationship between zygosity and clothing choices among male twins. Smith concluded that these results cast doubt on the validity of the equal environments assumption.

While most studies of the EEA have examined the post-natal environment, one early study focused on the pre-natal environment. Price (1950) considered how zygosity influences the prenatal environment and how zygosity-related differences in the prenatal environment are confounded with genetic effects on outcomes in twin studies. Some MZ twins share the same chorion[7] and some develop with separate chorions. By contrast, DZ twins always develop in separate chorions. These differences in pre-natal environment may matter because twins who share a chorion

[7] The chorion is a membrane that surrounds the embryo and other membranes.

compete for resources more than twins who develop within separate chorions. Price suggests that this difference in the prenatal environment makes some MZ twins less similar than we would expect them to be based on their identical genes. If it is true that similarity in prenatal environment serves to differentiate twins, estimates of genetic effects from twin studies could be biased downward. Alternatively, the effects of post-natal environmental similarity could be mitigated (at least to some extent) by the effects of pre-natal environmental similarity. However, the vast bulk of the literature on the EEA is concerned with the possibility that genetic effects are biased upward, not downward. Thus I focus on the former possibility and not the latter.

2.2 Early studies (1965–1979) examining whether genetic effects are confounded with environmental effects in twin studies

Prior to 1965, studies of the EEA found overwhelming evidence that the environments experienced by MZ twins were more similar than the environments experienced by DZ twins. However, as behavior geneticists have pointed out, simply finding that zygosity is related to environmental similarity is not enough to invalidate the EEA and undermine the basis of twin studies. In order to challenge the basis of twin studies, one must also find that environmental similarity is related to similarity in twin outcomes. If similarity in twin environments is related to similarity in twin outcomes, then the effects of zygosity on twin outcomes may be partially or fully mediated by environmental similarity.

If environmental similarity partially or fully mediates the effects of zygosity on twin outcomes, then the results of the twin study are rather ambiguous. Recall that twin studies are supposed to reveal the effects of genes on a particular trait. But it is possible that the genetic effects found by twin studies only partially reflect the effects of genes on a particular trait. The genetic effects found in twin studies may also reflect the influence of a special MZ twin environment that enhances similarity on a variety of traits.

Smith (1965) considered the possibility that estimates of heritability from twin studies are confounded by the effects of special twin environments. Twin environments may be characterized by intense co-socialization. Due to their genetic similarity, MZ twins may be comfortable with the sameness induced by intense co-socialization. DZ twins, however, are sometimes genetically quite different, and may resist expectations of similarity brought on by intense co-socialization. Intense co-socialization may in fact lead DZ twins to seek niches that further distinguish themselves from their co-twins. As a result, special twin environments could reduce DZ similarity relative to MZ similarity. In other words, special twin environments could violate EEA.

The importance of special twin environments was investigated by Vandenberg (1966). Vandenberg reasoned that if the differences between twins result from twins reacting to one another, differences between twins on various traits should be correlated. His analysis revealed an absence of significant correlations between differences in twins on a variety of traits. That is, twins who were different in one way were not more likely to be different in another way. For example, Vandenberg found little evidence that twin differences in test scores were explained by a general within-family environmental factor.

The EEA received additional support from studies in the late 1960s and 1970s. I review these studies below.

2.2.1 Scarr (1968)

By 1968, behavior geneticists acknowledged that the environments of MZ twins tended to be more similar than the environments of DZ twins. This finding was not by itself sufficient to reject the EEA. In order to reject EEA, one must show that similarity in outcomes is at least partly a function of similarity in environments.

Scarr (1968) evaluated the EEA in an innovative way by examining twins whose zygosity was misperceived by their parents. Unlike most twin studies, Scarr diagnosed zygosity via blood test, rather than via questionnaire. Since blood tests provide more accurate diagnosis of zygosity than questionnaires, Scarr was able to discern whether parents were mistaken in their perceptions more accurately than previous researchers who ascertained zygosity with survey data. In Scarr's sample, 17.4% of MZ twins were thought by their mothers to be dizygotic. About 31% of DZ pairs were thought to be monozygotic.

In some ways, examining cases of misperceived zygosity is an ideal test of the equal environments assumption. If MZ twins who are perceived as dizygotic are more similar to accurately perceived MZ twins than to accurately perceived DZ twins, then the EEA seems sound. If MZ twins who are perceived as dizygotic twins are more similar to accurately perceived DZ twins than to accurately perceived MZ twins, then the EEA may be unsound. The same logic applies when comparing the similarity of misperceived DZ twins to the similarity of correctly perceived twins.

Some of Scarr's results are reproduced in Table 3. Mothers evaluated their twins with respect to 300 adjectives, which were combined into 26 scales. Four of the scales appear in Table 3. Table 3 displays mean differences between co-twins for

Table 3: Mean differences between co-twins on various measures, reproduced from Scarr (1968)

	Correctly classified		Misclassified	
	MZ (n=19)	DZ (n=22)	MZ (n=4)	DZ (n=7)
Adjective check list scale 1	4.1	8.5	5.7	6.6
Adjective check list scale 2	5.6	15.8	7.3	9.7
Anxiety	4.7	12.0	4.0	5.9
Maturity scale	0.4	1.1	1.3	1.8

pairs whose zygosity was correctly classified by their parents and pairs who were incorrectly classified by their parents.

The tiny size of Scarr's sample precludes meaningful significance testing. But one can make inferences based on consistent trends in the data. Among correctly classified twins, MZ pairs are consistently more similar than are DZ pairs. Among misclassified twins, MZ pairs are also consistently more similar than DZ pairs, although the differences are narrower. Thus, the effect of true zygosity remains after controlling for perceptions of zygosity. If we look within (control for) categories of true zygosity, we find apparent effects of perceived zygosity as well. MZ twins mistaken for DZ twins are less similar than correctly classified MZ twins on three out of the four personality scales. (The exception is anxiety, with co-twin differences of 4.7 and 4.0 for correctly and incorrectly classified pairs, respectively.) DZ twins mistaken to be monozygotic are more similar than correctly classified DZ twins on three out of four personality characteristics. Overall, the data suggests that the effect of true zygosity is probably greater than the effect of perceived zygosity. However, the data leave open the possibility that perceptions matter. While the EEA is not eviscerated, it also does not emerge unscathed. Scarr (1968) is interpreted by later authors as offering unequivocal support for the EEA, but on close inspection, the evidence seems more ambiguous.

2.2.2 Matheny et al. (1976)

Matheny et al. (1976) examined violations of EEA arising from MZ twins' physical similarity. MZ twins may turn out more similar to each other than DZ twins do because MZ twins' physical similarity prompts similar treatment. Matheny et al. (1976) base their findings on a 1973 survey of parents of twins ages 3 $\frac{1}{2}$ to 13. The authors used a comprehensive scale of physical similarity, consisting of answers to questions about height, weight, facial appearance, hair color, eye color and complexion, in addition to answers to questions about whether the twins were "as alike as two peas in a pod" and whether parents or other family members ever confused the twins.

Surprisingly, similarity in appearance did not correlate with similarity in IQ, reading skills, or speech accuracy. It is unlikely that the absence of findings resulted from low reliability, at least in the case of IQ. IQ was measured in a rigorous way via the Stanford-Binet Intelligence Scale for younger children and via the Wechsler Intelligence Scale for Children (WISC) among slightly older children. In addition, Matheny et al. (1976) found few significant correlations between similarity in appearance and similarity in personality characteristics. The authors also found no significant correlations between similarity of dress and similarity of behavior within twin pairs. A weakness of the study is that the measures of similarity were based on reports from parents, and not from the twins themselves. The authors may have obtained a different result if their data included reports from the twins themselves or reports from others who observed the twins. Nevertheless, the results of Matheny et al. (1976) boost confidence in the validity of the equal environments assumption.

2.2.3 Landmark twin study of Loehlin and Nichols (1976)

Loehlin and Nichols (1976) may be the most comprehensive evaluation of the equal environments study to date. Loehlin and Nichols (L&N) gathered data on twins identified from lists of high school juniors who had taken the National Merit Scholar Qualifying Test (NMSQT) in 1962. The dataset was larger than most previous twin studies in the US, consisting of 514 MZ twins and 336 DZ twins. Hence, power was considerably higher in Loehlin and Nichols (1976) than in most previous twin studies.

Loehlin and Nichols (1976) evaluated the EEA in a number of ways. They examined whether mothers' perceptions of the zygosity of their twins influenced mothers' treatment of their twins, independent of twins' actual zygosity. Controlling for diagnosed zygosity, L&N found that the partial correlation between perceived zygosity and parental treatment of the twins was very low.[8]

[8] In the 1970s, it was not feasible to ascertain zygosity via biological assay through the mail. Instead, zygosity was predicted based upon reported degree of physical similarity between twins. Measurement error in Loehlin and Nichols (1976)'s measure of true zygosity was likely very low,

Using the NMSQT data, I reexamined L&N's finding that perceptions of zygosity do not matter when diagnosed zygosity is controlled.[9] Like L&N, I examined whether the parents' answers to six questions about similarity of rearing environment differed by *perceived* zygosity, after controlling for *diagnosed* zygosity. Parents responded to questions about whether the twins: played together or separately between ages 6 and 12, tended to spend their time together between ages 12 and 18, had the same teacher in school, slept in the same or different rooms, and whether the parents tried to treat the pair the same or differently. I combined the answers to six questions about environmental similarity into a composite measure that ranged between 11 and 23. Since reports of similar treatment were much more common than reports of differential treatment, the distribution of the composite treatment variable is skewed to the left.

The means on the composite measure of parents' reports of environmental similarity for each category of perceived and diagnosed zygosity are presented in Table 4. Higher values indicate higher levels of similarity in environment, according to parental reports. The differences in means are much greater between categories of diagnosed zygosity than between categories of perceived zygosity. A twoway ANOVA confirmed that the effect of perceived zygosity is not significant when controlling for diagnosed zygosity ($F = 0.45, p \approx 0.501$).

I examined the impact of perceived and diagnosed zygosity on *twins'* reports of environmental similarity. The results are presented in Table 5. The scale of environmental similarity is the sum of available twin reports on the following characteristics: similarity of clothing, same or different friends, longest period of separation from twin, proportion of time spent together, and proportion of possessions shared. Each of these measures was recoded so that high values indicate higher levels of shared experience. In the vast majority of cases, the data include reports on shared experience from both twins. These reports were averaged and then summed together to form a composite measure of environmental similarity. While the composite measure of parents' reports is skewed to the left, the composite measure of twins' reports is symmetrical. Twins appear more likely to acknowledge differences in their early environment than do their parents.

Table 5 presents averages on the composite measure of environmental similarity by both perceived and by diagnosed zygosity. The results of a twoway ANOVA reveal that the effect of perceived zygosity is significant when controlling for diagnosed zygosity ($F = 5.71, p \approx 0.02$). It appears that Joseph (2004) was correct in his

but not zero. Based on the results of previous validation studies, L&N estimate that their method miscategorizes between 5% and 10% of twins.

[9] John Loehlin generously made the NMSQT twin data available to me.

Table 4: Parental reports of environmental similarity, by perceived and diagnosed zygosity, NMSQT twins study

	Diagnosed zygosity	
Perceived zygosity	MZ	DZ
MZ	19.40	18.17
	390	52
DZ	19.36	17.80
	85	255
DK	20.00	19.55
	2	11

supposition that the use of parental reports masks the effect of perceived zygosity. To be sure, the effect of diagnosed zygosity is considerably larger than the effect of perceived zygosity.

Loehlin and Nichols (1976) presented additional evidence in support of the equal environments assumption. They calculated correlations relating the absolute difference between twins on a variety of traits to parental reports of environmental similarity. Traits included scales measuring various personality characteristics, and scales measuring vocational orientation.[10] Correlations were calculated separately for MZ and for DZ twins. Lower (more negative) correlations indicate that similarity in environments affects similarity in outcomes – in violation of the EEA.

The results from Chapter 5 of L&N are presented in Figure 2. Correlations are generally weak, and are rarely significant. In addition, there are about as many positive correlations as negative correlations. The average of all of the correlations is 0.02. Largely on the basis of this evidence, L&N concluded that the EEA was sound. However, L&N did not examine whether differences in traits correlated with *twins'* reports of environmental similarity. If Joseph (2004) was right in arguing that parents downplay environmental differences between their twin children, then parents' reports of environmental differences may be deceptively small.

I examined the correlations between absolute differences in co-twins' traits and

[10]The traits examined are identified in Figures 4–7. Since the present reanalysis of L&N is secondary to the main analysis in this paper, I do not review L&N's measures in depth.

Figure 2: Correlations between absolute differences on traits and parents' reports of environmental similarity, following Loehlin and Nichols (1976)

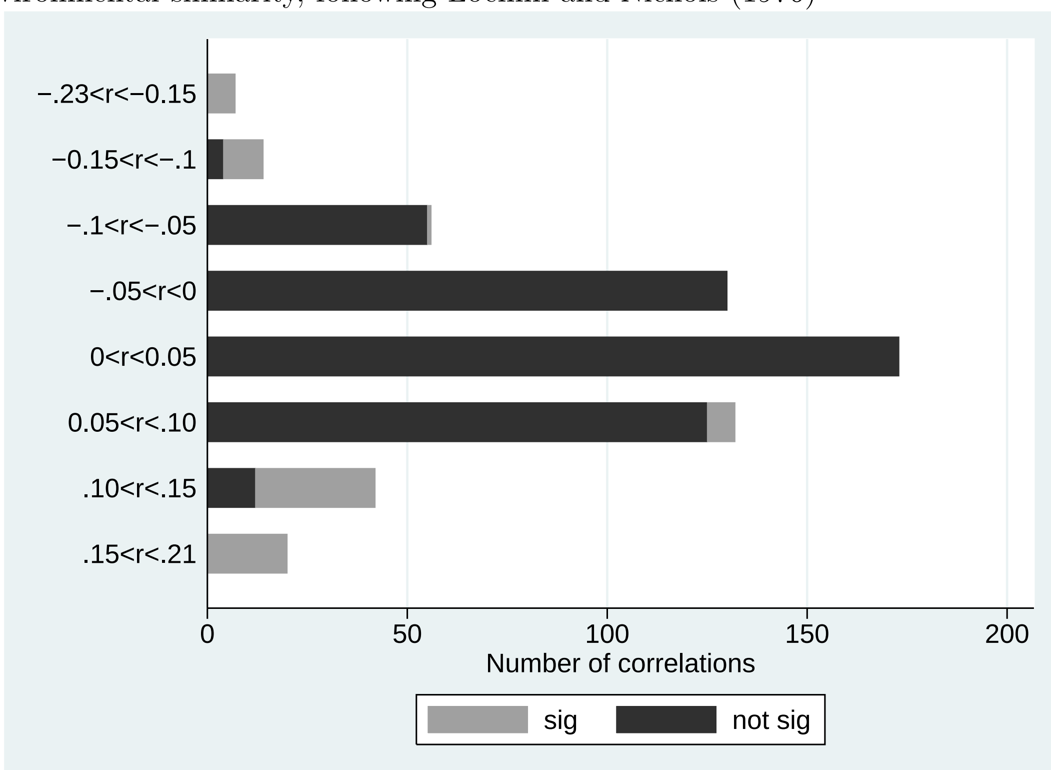


Table 5: Twin reports of environmental similarity, by perceived and diagnosed zygosity, NMSQT twins study

	Diagnosed zygosity	
Perceived zygosity	MZ	DZ
MZ	15.52	14.30
	411	53
DZ	15.11	13.18
	90	261
DK	19.50	15.59
	2	11

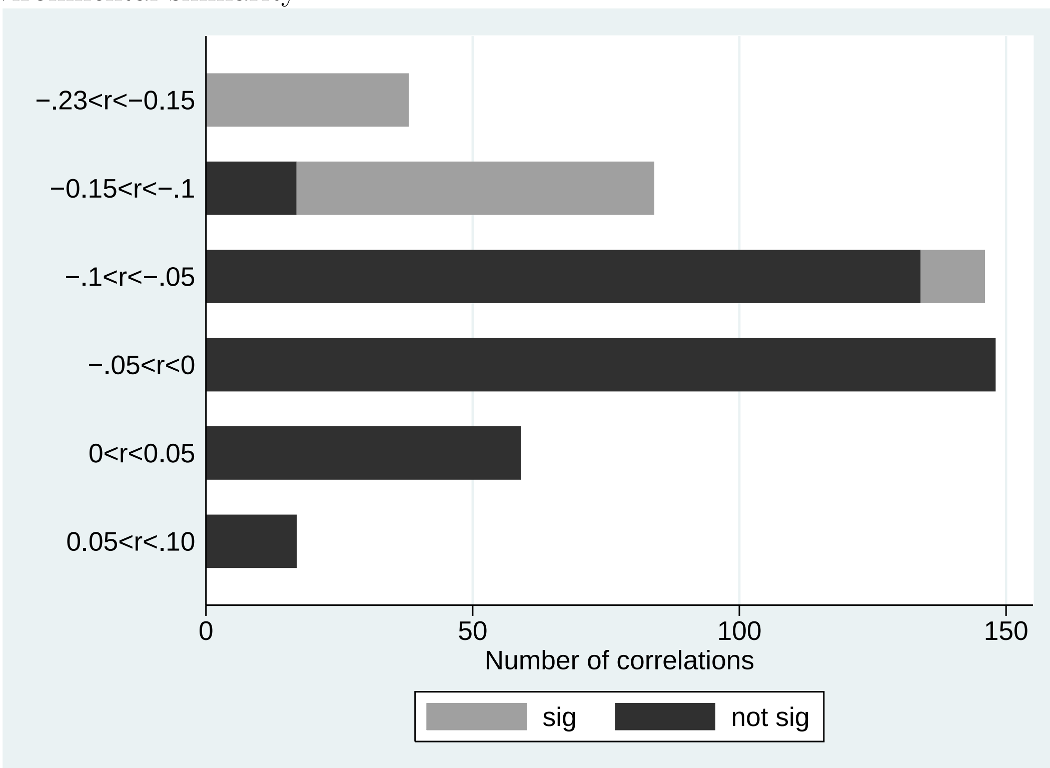
twins' own reports of the similarity of their environment. Here I looked at the same set of traits that formed the basis for the analysis presented in Figure 2. The results are presented in Figure 3. Only 14% of the correlations are significant. Most of the negative correlations are weak, although a substantial minority of the correlations are moderate in magnitude. Overall, the equal environments assumption is not fully supported, but the violations of that assumption are modest.

There is reason to believe that estimates of heritability from the NMSQT data might be reduced by controlling for twins' reports of environmental similarity. L&N did not test this possibility, since the negligible correlations between parents' reports of environmental similarity and trait similarity convinced them not to pursue the issue further. I found that more correlations between twins' reports of environmental similarity and trait similarity are significant than would be expected by chance. Based on this evidence, it makes sense to examine estimates of heritability for the traits examined by L&N with and without controls for environmental similarity.

I estimate heritability using DeFries-Fulker regression. Since I describe this method extensively in the Data and Methods section later on, I do not describe the method here. In order to control for environmental similarity, I employ propensity score matching. Since I describe how I conduct propensity score matching extensively in the Data and Methods section, I omit a discussion of matching here.

Figures 4 – 7 present estimates of the heritability of traits on matched and unmatched twins. Figure 4 displays heritability estimates for components of the Na-

Figure 3: Correlations between absolute differences on traits and twins' reports of environmental similarity



tional Merit Qualifying Test. In general, controlling for environmental similarity reduces heritability estimates of test scores very modestly. On average, the heritability estimates of the tests on matched data are only about 9% lower than the heritability estimates on unmatched data.

Figures 5 and 6 present estimates of heritability for various measures of personality. The differences in heritability estimates between matched and unmatched data are considerably larger for personality characteristics than for scholastic achievement tests. On average, controlling for environmental similarity reduces heritability estimates by about one third (37.6%). It does not seem that L&N were justified in ignoring environmental similarity when examining the influence of genes on personality.

Figure 7 shows estimates of heritability for vocational interests. Again, L&N argued in Chapter 5 of their book (1976) that environmental similarity was not a concern when estimating heritability. Figure 7 shows otherwise. In nearly all cases, heritability appears lower when estimated using matched data. To be sure, the differences between adjusted and unadjusted estimates are often not large enough to warrant wholesale revision of the central conclusion of behavior genetics. Using twins matched on environmental similarity, I still find substantial genetic effects. However, it appears that genetic effects may be exaggerated in the standard twin study. On average, estimates of heritability for all of L&N's traits examined here were about one third (33.9%) lower in matched than in unmatched twin pairs.

2.2.4 Lytton (1977)

Lytton (1977) employed a novel approach to examine the validity of the equal environments assumption. Unlike previous studies, Lytton had real-time observations of parent-twin interactions, in addition to survey responses from parents. Lytton analyzed the interactions of parents of 17 MZ and 29 DZ twin boys who were $2\frac{1}{2}$ years old.

While previous researchers focused on whether MZ twins are treated more similarly than are DZ twins, Lytton attempted to determine why. Are parents of MZ twins simply responding to the twins' genetic similarity, or are parents of MZ twins prompted by social expectations to treat their twins more similarly? To address this question, Lytton categorized his observations of parental behaviors toward their twins according to whether such behavior was prompted by the actions of their twins, or whether the behavior was initiated by parents themselves. If parents similar treatment is simply a response to similar twin behavior, then the equal environments assumption is (arguably) unproblematic. On the other hand, if similar treatment is

Figure 4: Heritability estimates for components of National Merit Qualifying Test: unmatched and matched twin pairs, NMSQT data

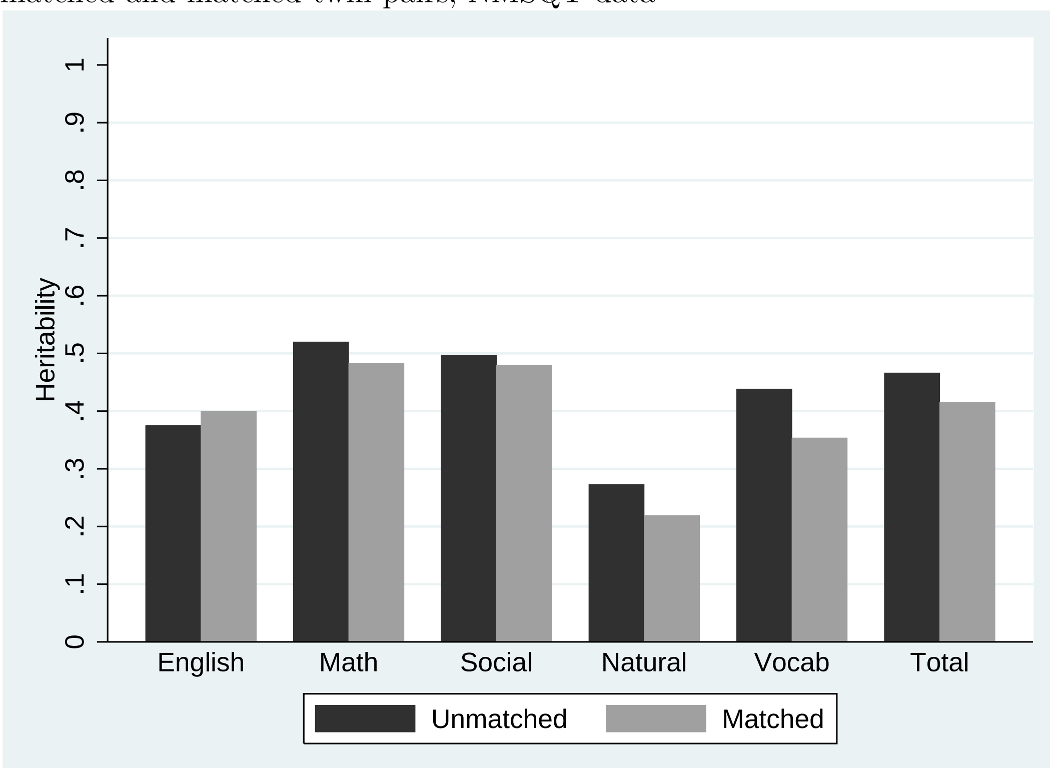


Figure 5: Heritability estimates for personality traits: unmatched and matched twin pairs, NMSQT data

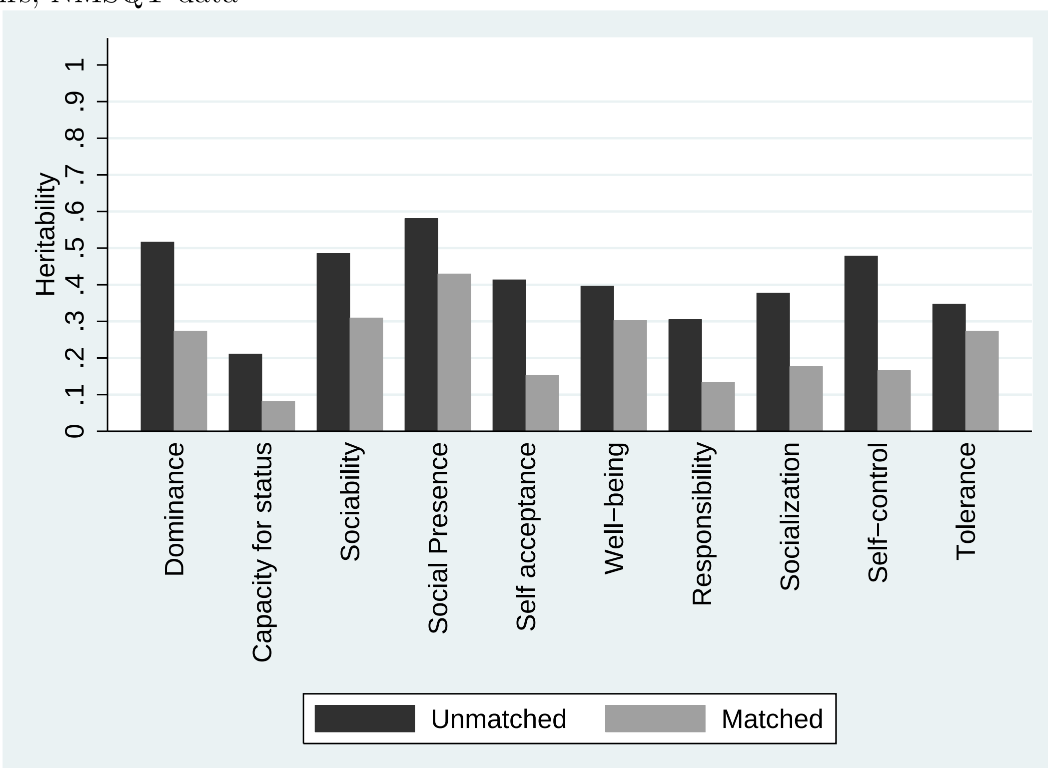


Figure 6: Heritability estimates for personality traits: unmatched and matched twin pairs, continued, NMSQT data

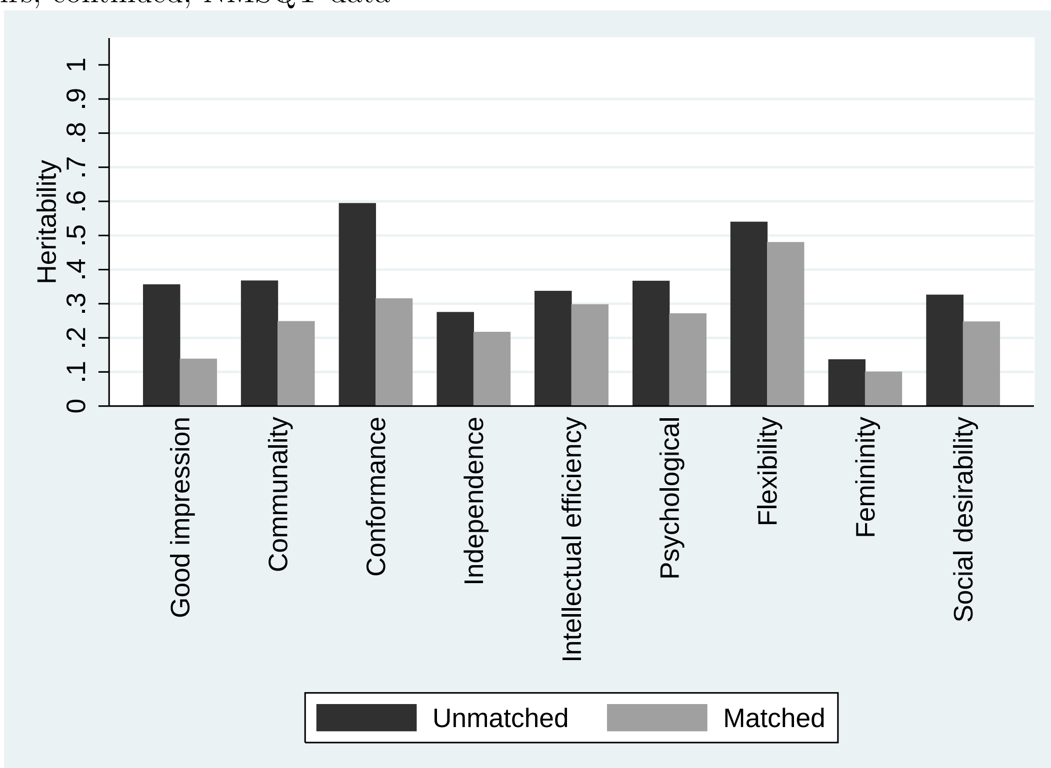
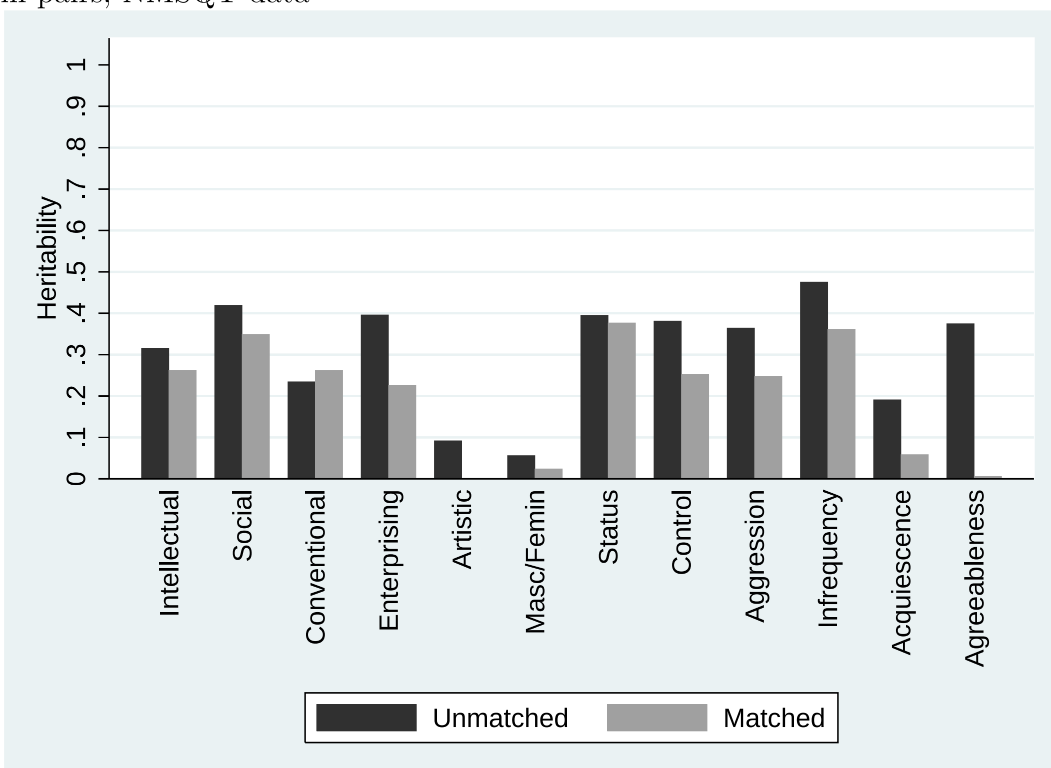


Figure 7: Heritability estimates for vocational interests: unmatched and matched twin pairs, NMSQT data



parent-initiated, then it may be prompted by social expectations about MZ twins, and the equal environments assumption may be unsound.

Based on an analysis of direct observations of mother-child interaction, Lytton concludes that parents "do not, of their own accord and systematically, institute more similar treatment for MZ than for DZ twins in general, and that the greater similarity of certain kinds of parent behavior towards MZ than towards DZ twins [where it exists] is likely to be a reaction to the MZ twins' own greater phenotypic likeness."

Lytton's interpretation of survey data collected from the mothers of the $2\frac{1}{2}$ year old twins is not persuasive. Mothers who acknowledged that they treated their twins differently reported that their twins seemed to have different needs. For Lytton, this fact appears to provide further evidence in support of the idea that mothers' actions are merely an outgrowth of MZ twins' phenotypic similarity. But how else would one expect parents to explain differences in the way they treat their children? How else could parents justify treating their children differently, except to say that they were doing so in reaction to their children's behavior? Treating children differently for any other reason would appear unfair.

I do not think that the validity of the equal environments assumption hinges on whether treatment similarity is a function of twins' phenotypic similarity, or whether treatment similarity is a function of social expectations. Even if MZ twins' phenotypic similarity is solely responsible for treatment similarity, the equal environments assumption is not necessarily sound. As I have argued before, if similarity in environments is a *necessary* cause of similarity in outcomes, then the results of twin studies may not generalize to the non-twin population.

2.2.5 Scarr and Carter-Saltzman (1979)

Scarr and Carter-Saltzman (1979) expanded on Scarr (1968)'s previous analysis examining twins who misperceived their zygosity. In a sample of 400 pairs of twins, ages 10 to 16, the authors ascertained true zygosity by blood test, and compared these results with twins' perceptions of zygosity. Surprisingly, only 60% of twin pairs were correct about their zygosity. In the other 40% of cases, one or both twins reported incorrect beliefs about their zygosity.

Scarr and Carter-Saltzman (S&C) ran regressions predicting absolute differences between twins' cognitive ability, personality, and physical characteristics. Regressors included actual and perceived zygosity, as well as similarity in appearance. S&C argue that their findings support the equal environments assumption.

But their analyses are set up in a way that makes it unlikely they will detect violations of the equal environments assumption. S&C examine the effects of perceived similarity when controlling for actual zygosity. However, S&C do not examine whether the effects of actual zygosity decline significantly when controls for perceived zygosity are introduced. Thus, it is unclear to what extent the effects of true zygosity are confounded with the effects of perceived zygosity. To be sure, the effect of true zygosity is likely to be much more powerful than the effect of perceived zygosity. When controlling for true zygosity, S&C found that the effects of perceived zygosity, as well as most other measures of perceived similarity, were not significant. The reverse is unlikely to be true.

But S&C did not give the effect of perceived zygosity a fair chance. S&C include five measures of perceived zygosity in their regressions *simultaneously*: (1) subjective judgment of zygosity by survey administrators, and twin-reported measures of: (2) dressing alike, (3) looking alike, (4) cases of mistaken identity, and finally, (5) twins' perception of zygosity. The correlations between these variables, as reported by S&C, range between .151 and .554. The alpha coefficient for these five variables is 0.76. Given that all of these measures were included in a regression run on a case base of about 100 twin pairs, it is unsurprising that S&C found few significant effects. There is likely little power to detect independent effects of 5 different measures of perceived zygosity due to multicollinearity. Multicollinearity may be especially acute because S&C included seven measures of physical similarity such as height and weight in these regressions. These variables are likely to be at least moderately correlated with the measures of perceived zygosity, and there is likely precious little variance left to tell us anything. Thus, the results of S&C's analyses are not as illuminating as they could be and they are not very helpful in detecting violations of the equal environments assumption.

It does not appear that Scarr and Carter-Saltzman (1979) tested the equal environments assumption in a rigorous way. S&C regressed absolute differences in twins' cognitive tests on measures of perceived similarity on a sample that included both MZ and DZ twins. The results of these regressions were less clearly favorable toward EEA than S&C would have us believe. I reproduce the regression results in Table 6.

Table 6 is reproduced from Table VI on p. 538 of S&C. Each column in the table displays coefficients from regressions predicting twin differences in cognitive skill. The dependent variables in these regressions are absolute differences between co-twins on the following tests: Raven Progressive Matrices (Raven), Peabody Picture Vocabulary Test (PPVT), Columbia Mental Maturity Scale (Columbia), and the Revised Visual Retention Test (Benton). The independent variables include three groups of variables: five measures of perceived similarity, seven anthropometric measures, and two genetic measures. Measures of perceived similarity include a

Table 6: Regressions of absolute twin differences in cognitive test scores on measures of perceived, physical, and genetic differences, n=226, reproduced from Scarr and Carter-Saltzman (1979)

	Test			
Co-twin differences	Raven	PPVT	Columbia	Benton
Perceptions				
Eight raters	-0.23	-0.02	-0.05	0.03
Dress alike	0.11	-0.08	-0.02	0.02
Look alike	0.00	-0.02	0.03	0.15
Mistaken	-0.06	0.09	-0.18	-0.19
Twin zygosity	-0.05	0.10	0.03	0.03
Physical				
Skeletal age	-0.03	-0.16	0.00	0.04
Upper arm circumference	0.04	-0.04	-0.05	-0.08
Triceps skin fold thickness (TSFT)	0.01	0.05	-0.01	-0.03
Sitting height	0.01	0.11	-0.03	0.09
Weight	-0.09	0.08	0.09	0.03
Skin reflectance (RFH)	0.07	-0.04	0.05	-0.04
Stature	-0.01	0.11	0.04	0.05
Genetic				
Number of blood group differences	0.44	0.37	0.08	0.21
Dizygosity	0.01	0.34	0.04	-0.09
R^2	0.11	0.09	0.04	0.06

scale of appearance similarity ratings from eight independent judges. The other four measures of perceived similarity, including the measure of perceived zygosity, come from the twins themselves. Measures of physical similarity were all based on direct measurement. Finally, the genetic measures include a measure of actual zygosity, in addition to a measure of blood group similarity on which the measure of actual zygosity is based.

Due to the inclusion of very similar measures, multicollinearity is most likely a problem here, as it was in the other analyses conducted by S&C. For this reason, the regressions do not permit adjudication between the effects of actual versus perceived zygosity. Had S&C combined similar measures into scales, the regression results could have shed more light onto the validity of the equal environments assumption. As it stands, neither the effect of perceived zygosity nor the effect of actual zygosity are significant in most regressions. Yet we cannot expect measures of actual or perceived zygosity to be significant, since measures closely similar to these were also included. Had the regressions been run differently, it is possible that we would find significant effects of both real and perceived zygosity. In addition, it would have been helpful for S&C to show the effects of actual zygosity on similarity in cognitive test scores with and without controlling for perceived zygosity. The effect of real zygosity would likely decline, yet remain significant, when a control for perceived zygosity was included in the model.

S&C also present a multivariate ANOVA in which they present absolute differences in cognitive and personality tests by true zygosity, perceived zygosity, and the interaction between true and perceived zygosity. Again, the results are not as resolutely in favor of the equal environments assumption as the authors suggest. It is true that perceived zygosity does not have a significant effect on any of the cognitive tests, nor any of the personality tests. Nonetheless, when controlling for perceived zygosity, the effects of true zygosity are only significant for two out of four cognitive tests, and in one of two of the personality tests. Specifically, true zygosity has significant effects on the similarity of scores on the Raven's Progressive Matrices, and has a significant effect on the similarity of scores on the Revised Visual Retention Test. Yet for both of these tests, there are also significant interactions between true and perceived zygosity. The interpretation for the interaction effects is unclear. Consider Tables 7 and 8, which are extracted from Table VIII of Scarr and Carter-Saltzman (1979).

If similarity is purely a result of genes, not social forces, then average absolute differences in test scores should differ by true zygosity only. In this case, the figures be approximately the same with rows, and different across rows in Tables 7 and 8. If similarity is purely a result of social forces, not genes, then average absolute

Table 7: Average absolute differences between co-twins on Raven's Progressive Matrices, from Scarr and Carter-Saltzman (1979)

		Perceived zygosity MZ	DZ
True zygosity	MZ	0.66	0.80
	DZ	1.11	0.82

Table 8: Average absolute differences between co-twins on Revised Visual Retention Test, from Scarr and Carter-Saltzman (1979)

		Perceived zygosity MZ	DZ
True zygosity	MZ	0.61	0.97
	DZ	0.89	0.83

differences in test scores should differ by perceived zygosity only. In this case, the figures in Tables 7 and 8 should be approximately the same within the columns, and different between columns. If we compare the average absolute differences in test scores across rows and columns in Tables 7 and 8, the story is not so clear. For the results with Raven's Progressive Matrices in Table 7, we see a result consonant with the genetic view in the first column (.66 versus 1.11), but a result less consonant with the genetic view in the second column (.80 versus .82). Among twins perceived as MZ, real zygosity seems to have a substantial effect on test scores. Yet among twins perceived as DZ, real zygosity appears to little or no effect on test scores. Looking within rows yields a similarly confusing picture. Among true MZ twins, perceived zygosity appears to make a difference in terms of test score similarity. Among true DZ twins, the effect of perceived zygosity is in the opposite direction from what one would predict. So it is a messy picture, one that might arise from randomness owing to the relatively small sample sizes. The results for the Revised Visual Retention Test are no clearer. The effect of actual zygosity may be stronger than the effect of perceived zygosity, but these results are not sufficiently clear cut to warrant the conclusion that similarity in treatment can be ignored in twin studies.

2.3 Recent studies of the equal environments assumption

Since I have discussed most of the issues that arise in the literature examining the EEA, I provide a condensed summary of the results of the most recent research. Table 9 introduces all of the studies I could find which test the equal environments assumption. Most of the studies were based on probability samples of substantial size. In Table 10, I summarize the measures of environmental similarity used in each study.

The measures of environmental similarity in these studies can be classified into five general categories:

- similarity of appearance: Hettema et al. (1995), Klump et al. (2000), Horwitz et al. (2003)
- similarity of childhood environment: Morris-Yates et al. (1990), Kendler and Gardner (1998), Borkenau et al. (2002), Cronk et al. (2002), Horwitz et al. (2003)
- cases of mistaken zygosity: Kendler et al. (1993), Cronk et al. (2002), Gunderson et al. (2006)

- frequency of contact over the life course: Rose et al. (1988), Eriksson et al. (2006)
- bonding or attachment: Kendler and Gardner (1998), Penninkilampi-Kerola and Moilanen (2005)

Collectively, the studies cover a broad range of environmental measures; however, no study controls for all of the ways in which environmental similarity differs by zygosity. The present study will control for environmental similarity in a more comprehensive way than all previous studies have done.

Table 11 provides a list of the outcomes for each of the thirteen studies since the late 1980s that tested the equal environments assumption. Each of these studies examined whether similarity in environments between co-twins was related to twin similarities on the specified outcomes, controlling for zygosity.

Most of the outcomes examined by studies of the EEA are either personality characteristics (e.g. extraversion, neuroticism, agreeableness, conscientiousness, and openness to experience) or psychological disorders (e.g. depression, phobias, generalized anxiety disorder, etc.) There has little research testing the EEA with respect to outcomes that interest sociologists, such as educational attainment, occupational attainment, religiosity and political attitudes. Researchers estimating the heritability of education and political attitudes did not attempt to control for environmental similarity (Nielsen, 2006; Alford et al., 2005). The present study estimates the heritability of these and many other traits while controlling for environmental similarity. The present study examines the validity of the EEA for a larger variety of outcomes than any previous study. Examining a large variety of outcomes on a single dataset permits me to make more robust conclusions about the extent to which the validity of EEA varies according to outcome.

Table 12 presents information about the method used by each study, and also provides a rough summary of the number of violations of EEA discovered. More precisely, the right-most column of Table 12 provides a tally of the hypotheses in which EEA was violated out of the total number of hypotheses testing EEA.

It is immediately apparent that very few studies reveal problems with the EEA. Indeed, the vast majority of authors interpreted their findings as confirmation that the equal environments assumption was well-founded. Although the conclusions of these studies are consistent, recall that the scope of the findings is limited mostly to personality characteristics and psychological disorders. In addition, none of the studies examined here used a comprehensive set of control variables for environmental similarity.

The findings of Penninkilampi-Kerola and Moilanen (2005), which could not be characterized in a simple manner in Table 12, suggest that previous examinations of EEA are missing a key control variable. Adjusting for a single dichotomous measure of co-twin self-rated dependence on each other, Penninkilampi-Kerola and Moilanen (2005) found significant reductions in estimates of heritability of alcohol use and abuse. Aside from Penninkilampi-Kerola and Moilanen (2005), only one other recent study used a measure of twin attachment (Kendler and Gardner, 1998). The present study improves upon previous literature by controlling for multiple measures of twin attachment.

Table 9: Recent studies testing the equal environments assumption

Study	Sample	# MZ pairs	# DZ pairs
Rose et al. (1988)	Finnish Twin Cohort Study	2,320	4,824
Morris-Yates et al. (1990)	Australian Twin Registry	186	157
Kendler et al. (1993)	Females from Virginia Twin Registry	1,082	790
Hettema et al. (1995)	Females from Virginia Twin Registry	525	349
Kendler and Gardner (1998)	Females from Virginia Twin Registry	525	349
Klump et al. (2000)	Females in the Minnesota Twin Family Study	186	100
Borkenau et al. (2002)	Convenience sample in Germany	525	268
Cronk et al. (2002)	Population-based sample of females from Missouri	1,093	855
Horwitz et al. (2003)	National Longitudinal Study of Adolescent Health	230	187
Penninkilampi-Kerola and Moilanen (2005)	Finnish Twin Cohort Study	1,655	1,707
Derks et al. (2005)	Netherlands Twin Register	529	1,005
Eriksson et al. (2006)	Swedish Young Male Twins Study	359	232
Gunderson et al. (2006)	Kaiser Permanente Twin Registry	205	145

Table 10: Recent studies: measures of environmental similarity

Study	Measures
Rose et al. (1988)	length of life with co-twin, frequency of contact
Morris-Yates et al. (1990)	parental treatment, dress, share bedroom, same classes, play together (6-11), spent time together (12-16)
Kendler et al. (1993)	self-reported zygosity versus assigned zygosity
Hettema et al. (1995)	physical similarity ratings of photographs
Kendler and Gardner (1998)	7 measures of childhood environment, emotional closeness in childhood, emphasis on similarity by twins & others
Klump et al. (2000)	physical similarity: physical measurements and judgments of photos
Borkenau et al. (2002)	similarity of childhood experiences, self- and peer ratings
Cronk et al. (2002)	Mothers' perceptions of zygosity and similarity of childhood experience
Horwitz et al. (2003)	similarity in attractiveness, amount of time spent together, same friends, similarity in the # friends who drink
Penninkilampi-Kerola and Moilanen (2005)	Self-reported co-twin dependence
Derks et al. (2005)	model-based: allowed the correlation between shared environments to be estimated for DZ twins
Eriksson et al. (2006)	frequency of contact
Gunderson et al. (2006)	self-reported zygosity versus diagnosed zygosity

Table 11: Recent studies: outcome measures

Study	Outcomes
Rose et al. (1988)	extraversion, neuroticism
Morris-Yates et al. (1990)	neuroticism, anxiety, depression
Kendler et al. (1993)	major depression, generalized anxiety disorder, phobia, bulimia, alcoholism
Hettema et al. (1995)	major depression, generalized anxiety disorder, phobia, bulimia, alcoholism
Kendler and Gardner (1998)	major depression, generalized anxiety disorder, panic disorder, phobias, bulimia, alcoholism, smoking
Klump et al. (2000)	eating disorder inventory: total, body disatisfaction, compensatory behavior, binge eating, weight preoccupation
Borkenau et al. (2002)	neuroticism, extraversion, openness to experience, agreeable, conscientiousness
Cronk et al. (2002)	separation anxiety disorder, ADHD, oppositional defiant disorder, and conduct disorder
Horwitz et al. (2003)	Body mass index, depression, trying alcohol, frequency of drinking, frequency of binge drinking
Penninkilampi-Kerola and Moilanen (2005)	drinking frequency, intoxication frequency
Derks et al. (2005)	Spatial ability & aggression, direct & indirect
Eriksson et al. (2006)	physical activity
Gunderson et al. (2006)	biometric measures, diet

Table 12: Recent studies: methods and findings

Study	Method	Violations found of those tested
Rose et al. (1988)	regression	3 of 4
Morris-Yates et al. (1990)	partial correlations	3 of 12
Kendler et al. (1993)	SEM with contingency tables	0 of 5
Hettema et al. (1995)	SEM with contingency tables	1 of 5
Kendler and Gardner (1998)	logistic regression	2 of 48
Klump et al. (2000)	ANOVA, MANOVA, correlations	0 of 4
Borkenau et al. (2002)	regressions	3 of 30
Cronk et al. (2002)	multigroup SEM	0 of 4
Horwitz et al. (2003)	regression	2 or 3 of 6
Penninkilampi-Kerola and Moilanen (2005)	SEM with contingency tables	NA
Derks et al. (2005)	SEM	0 of 2
Eriksson et al. (2006)	SEM with contingency tables	0 of 4
Gunderson et al. (2006)	correlations by correct & mistaken zygosity	1 of 9

2.4 Studies of adopted children and studies of twins reared apart

Defenders of the equal environments assumption cite research on adopted children and research on twins reared apart as corroborating evidence for the classic twin study. Adoption studies and studies of twins reared apart are not subject to the same biases as the classical twin study. If heritability estimates from classic twin studies were biased upward due to violations of the EEA, then these estimates of heritability would be higher than the estimates of heritability based on other methods. In fact, the results of adoption studies 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, 1994, 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.

There are only a few major 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, 1994, p. 72). The most prominent study of twins reared apart in the US revealed moderate to high correlations on a variety of characteristics and behaviors (Bouchard et al., 1990). According to Bouchard et al. (1990), the correlations on personality between MZ twins reared apart were almost as high as the correlations on personality between MZ twins reared together. Moreover, estimates of the heritability of IQ from the two major studies of twins reared apart – Bouchard et al. (1990) and Pedersen et al. (1992) – are higher than many estimates of heritability from classic twin studies.

But studies of adopted children are not unassailable. Since parents must go through a screening process before they are permitted to adopt a child, the variation in adoptive environments is likely 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, estimates of heritability from adoption studies may not apply to the general population. Like the results of classic twin studies, the results of adoption studies may be biased.

There are also some concerns about the results of Bouchard et al. (1990), the most comprehensive study of twins reared apart in the US. Not surprisingly, 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 (2004) 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. So the sample of Bouchard et al. (1990) may be compromised by selection bias. Joseph (2004) also argues that it is impossible to evaluate or replicate Bouchard et al. (1990) since the authors have not been willing to release their data – even now, several decades after publication.

In general, the criticisms of behavior genetic studies are far from fatal, but these criticisms may have some merit. There is a possibility that estimates of heritability from the three types of behavior genetic studies – adoption, twins reared apart, and the classic twin study – are biased upward for different reasons. Thus, it makes sense to evaluate the possibility that the classical twin study is biased due to violations of the equal environments assumption.

3 Data

Most twin studies estimate heritability without controlling for environmental similarity. The few twin studies that do control for environmental similarity do so in a limited way with one or two variables. In this study, I examine how heritability estimates are affected by a wide variety of measures of environmental similarity. In an attempt to provide a comprehensive look at the equal environments assumption, I examine the heritabilities of a large number of outcomes, including those of greatest interest to psychologists as well as those of more interest to sociologists. The equal environments assumption is evaluated using several different methods in order to ensure that conclusions are not artifacts of any particular method.

I analyze data from the twin component of the Midlife Development in the United States (MIDUS) survey (Brim et al., 2008). 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. MIDUS is an apt choice for these analyses for several reasons. A larger and more diverse set of questions about environmental similarity were asked of twins in MIDUS than in other contemporary survey of twins. In addition, MIDUS contains a large number of questions on a variety of topics, including some of the most frequently examined in other twin studies, as well as other questions of interest to sociologists. Using MIDUS, I am able to examine how well the equal environments assumption holds up in analyses of outcomes of sociological interest, as well as how well the assumption holds up in analyses of outcomes for which genetic influence is less controversial, such as health and height.

MIDUS respondents were chosen via random digit dialing of the English-speaking non-institutionalized population in the lower United States ages 25-74 in 1995. As such, MIDUS is one of the very few samples of twins that are nationally representative. Since twins comprise only about 2% of the population, collecting a nationally representative sample of twins is an assiduous task. As a part of larger survey efforts, MIDUS phone interviewers screened about 50,000 households for the presence of a twin. About 15% of respondents reported a twin present in the family. Of those who reported a twin in the family, 60% agreed to provide contact information for the twin to interviewers.

The MIDUS survey instruments were substantial. Twin respondents were asked to participate in 30-minute phone interviews and were sent two 45-page self-administered questionnaires, as well as an additional short questionnaire asking for twin-specific information. The response rate for the phone interview was 60%. Among those who participated in the phone interview, 92% sent back self-administered questionnaires. The overall response was 60% x 92% = 55.2%. With regard to demographic composition, the twin sample is 45% male, and has a mean age of 45.

Attempts were made to interview MIDUS respondents again between 2004 and 2006. The response rate for twins in the follow-up survey wave, adjusted for mortality, was 82%. The second wave, like the first, consisted of both phone interviews and self-administered mail-back questionnaires. Many of the questions that were asked in wave one were asked again in wave two. As I will explain in more detail later, I used data from the second wave as a check on the reliability of heritability estimates from the first wave. However, the bulk of my analyses were conducted using data from the first wave only, since attrition substantially reduced sample size and statistical power at the second wave.

The sample of twins can be divided into three groups by zygosity and gender: monozygotic, same-sex dizygotic, and different-sex dizygotic. In the present analyses, monozygotic twins are compared with same-sex dizygotic twins. Different-sex dizygotic twin pairs were excluded because opposite sex pairs are not directly comparable to monozygotic pairs, who are always of the same gender.

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. Since I am going to examine the similarity of twins to each other, analysis is not possible when I have information from only one twin. Thus, forty twins whose co-twin did not respond to the survey were also deleted from the sample. After these deletions, the sample included 345 pairs of monozygotic twins, and 312 pairs of same-sex dizygotic twins.

3.1 Measures

The measures used in this study can be categorized into four groups: (1) fifty-nine outcome measures; (2) eight measures of environmental similarity; (3) a single measure of zygosity; and (4) several additional control variables, gender and age. Below, I describe how each of these measures was constructed. Given the large number of variables, I economize on space by providing much of the information about variable construction in tables.

3.1.1 Outcome measures

Since the purpose of this project is to evaluate the equal environments assumption in a comprehensive way, I examine a large variety of outcomes. Outcome measures were classified into the following thirteen categories: physical measurements, health, health-related behaviors, mental disorders, personality characteristics, self-actualization, well-being, sociability, marriage and family, social class, socio-political attitudes, altruism, and religion. The categorization scheme was developed based on the kinds of variables available in the dataset. Within each category, there are two or more outcome measures, most of which are scales constructed from multiple items, and a few of which are single items.

Below are thirteen tables containing information about the dependent variables in each category. Unless otherwise indicated, variables were constructed from answers to the self-administered questionnaires. Many of the scales used were constructed by MIDUS investigators. These scales are indicated with a superscript a. The construction of these scales is discussed in detail in the Scales Index component of the documentation for MIDUS archived at the Inter-University Consortium for Political and Social Research (Brim et al., 2008).

Listed in the rightmost column is the number of pairs of twins used in the analysis. The calculation of means and standard deviations (and subsequent analyses) is slightly complicated by the nature of twin data. For each outcome, we have two variables, one for each of the two twins. For each twin pair, the decision about which twin is recorded in variable one, and which twin is recorded in variable two, is arbitrary. This is a problem because calculations will vary depending on arbitrary reassignments of twin pairs to variables.

Table 13: Descriptive statistics for physical attributes

Outcome	Coding	Mean	StDev	Min	Max
Height	1 item, measured in inches	66.708	4.173	54.9	98.5
Weight	1 item, measured in pounds	165.149	36.108	95.5	364
Waist-to-hip ratio	Based on two items	0.875	0.096	.67	1.36
Body Mass Index (BMI)	Based on two items	26.261	5.062	15	57

Analysts have dealt with this problem in a number of ways. If there is reliable information about the birth order of the twins, then the twins can be assigned to variables in a less arbitrary manner. Although twins were asked about birth order in MIDUS, many of them could not recall this information, and so it was not feasible to order twins according to when they were born. Another possibility is that analyses could be run on every combination of twin orderings, and then averaged to obtain a final result. This solution is unwieldy because there are 2^n combinations of twin pairs for n observations. Running the present analyses 2^{600} times would be a daunting task even with modern computing.

A feasible solution is to enter each twin pair twice, with each twin taking turns as 'twin one' and 'twin two.' The double-entered data thus provide the basis for most analysis. Where necessary, degrees of freedom are set to equal the original number of twin pairs. Double-entry as become a standard practice in twin studies, and it is the approach that was adopted here. Means and standard deviations listed in the tables below were computed on the double-entered data.

Table 13 contains descriptive statistics for physical measurement outcomes. It is unlikely that the heritability of physical measurements will be much affected by controls for environmental similarity. The idea that height is largely driven by genetics in the developed world is uncontroversial. Few would argue that heritability estimates of height from twin studies are confounded with environmental influences. Thus, if heritability estimates of height are substantially affected by controls for environmental similarity, one might argue that we are controlling away true genetic effects. On the other hand, if genetic effects on uncontroversial outcomes like height

Table 14: Descriptive statistics for measures of Health

Outcome	**Coding**	**Mean**	**StDev**	**Min**	**Max**
Self-reported health	Sum of 2 items from phone interview	7.444	1.594	2	10
Number of chronic health conditions treated in the last 12 months	Sum of conditions that respondent checked off on a 29-item list[a]	2.182	2.367	0	21
Symptoms of illness	Mean of 9 items, each indicating the frequency of particular symptoms[a]	8.965	7.099	0	40
Ever diagnosed with cancer	1 item from phone survey. 1 if yes, 0 if no.	0.057	0.232	0	1
Any heart problems	Based on 2 items from the phone survey. 1 if yes, 0 if no.	0.136	0.343	0	1
Sensitivity to pain	Mean of 5 items[a]	2.485	0.669	-4	4

are immune to adjustments for environmental similarity, we can be confident that we are justified in subjecting twin studies to greater scrutiny in this way. Meanwhile, high heritability estimates of obesity have been disputed recently on methodological grounds (Martin, 2008).

Table 14 below contains descriptive statistics for health outcomes. Since the sample is relatively young, the vast majority of respondents have not had cancer or heart disease. Given the relatively young sample, the variance in health outcomes is relatively low, and so the power to detect effects of genes or environment will be limited for these outcomes.

Table 15 contains descriptives statistics for health-related behaviors. As noted in the Literature Review above, the equal environments assumption has been examined with respect to the heritability of smoking (Kendler and Gardner, 1998). So this will be an opportunity to replicate previous tests of EEA with regard to smoking frequency.

Table 15: Descriptive statistics for measures of Health-Related Behaviors

Outcome	**Coding**	**Mean**	**StDev**	**Min**	**Max**
Frequency of vigorous exercise in a month	Scale of 2 items[a]	6.294	5.207	0	13.5
Frequency of moderate exercise in a month	Scale of 2 items[a]	9.539	4.669	0	13.5
Respondent ever smoked.	Based on 1 item from phone survey. 1 if yes, 0 if no.	0.756	0.429	0	1
Number of cigarettes smoked per day during the year when smoked most heavily	Based on 3 items from phone survey; equals zero if person never smoked	11.053	15.669	0	97
Respondent started drinking regularly before age 21	Based on 2 items from phone survey. 1 if yes, 0 if no.	0.138	0.345	0	1

Table 16: Descriptive statistics for measures of Mental Disorders

Outcome	Coding	Mean	StDev	Min	Max
Depression scale 1 sadness	Number of yes responses to 7 items[a]	0.608	1.745	0	7
Depression scale 2 lost interest in life	Number of yes responses to 6 items[a]	0.141	0.772	0	6
Depression diagnosis	Based on depression scale 2 above, and 2 other items[a]	0.127	0.333	0	1
Generalized Anxiety Disorder	Number of yes responses to 10 items[a]	0.152	0.896	0	10
Panic attacks	Number of yes responses to 6 items[a]	0.364	1.099	0	6

Table 16 contains descriptive statistics for measures of mental disorders recognized by the American Psychiatric Association and catalogued in the Diagnostic Statistical Manual of Mental Disorders Association (2000). All of these scales were constructed by MIDUS investigators.

Table 17 contains descriptive statistics about the 'Big Five' personality characteristics in wide use by psychologists (McCrae and John, 1992). Each of the five traits listed – agreeableness, openness to experience, extraversion, conscientiousness and neuroticism – is believed to encompass a larger array of related traits. Personality psychologists have argued that self-reports on a large number of personality-related questions can be factor analyzed into these 'Big Five' attributes. Twin studies have found heritability estimates for each of the Big Five attributes on the order of about 0.4 (Rowe, 1994).

Table 18 contains descriptive statistics about measures relating to self-actualization and efficacy. These measures may be of interest to sociologists, as a sense of self-efficacy may predict status attainment.

Table 19 contains descriptive statistics for measures of well-being. These scales essentially measure the same construct as the mental disorder scales. Measures of

Table 17: Descriptive statistics for measures of Personality

Outcome	Coding	Mean	StDev	Min	Max
Agreeableness	Mean of 5 items [a]	3.527	0.466	1.2	4
Extraversion	Mean of 5 items [a]	3.219	0.559	1.4	4
Neuroticism	Mean of 4 items [a]	2.245	0.671	1	4
Conscientiousness	Mean of 4 items [a]	3.447	0.438	1.75	4
Openness to experience	Mean of 7 items [a]	2.969	0.524	1	4

Table 18: Descriptive statistics for measures of Self-Actualization

Outcome	Coding	Mean	StDev	Min	Max
Locus of control	Mean of 8 items a , labeled as perceived constraints by MIDUS investigators	2.557	1.225	1	6.63
Personal mastery	Mean of 4 items [a]	5.877	0.977	1	7
Sense of purpose	Mean of 3 items [a]	16.645	3.588	4	21
Autonomy	Mean of 3 items [a]	16.324	3.327	4	21
Present versus future time orientation scale 1	Mean of 3 items [a]	3.140	0.689	1	4
Present versus future time orientation scale 2	Mean of 3 items [a]	2.298	0.684	1	4

Table 19: Descriptive statistics for measures of Well-Being

Outcome	**Coding**	**Mean**	**StDev**	**Min**	**Max**
Self-esteem	Sum of 3 items, labeled as self-acceptance by MIDUS investigators[a]	16.902	3.462	3	21
Satisfaction	Sum of 3 items	9.975	2.336	1	12
Resilience	Mean of 5 items, labeled as persistence by MIDUS investigators[a]	3.245	0.534	1.4	4
Negative affect	Mean of 6 items [a]	1.504	0.617	1	5
Positive affect	Mean of 6 items [a]	3.468	0.714	1	5
Self-rated mental health	1 item	3.888	0.898	1	5

mental disorders are limited in their capacity to capture much of the range in the distribution of well-being; these measures should compensate.

Table 20 contains descriptive statistics for two measures of sociability. These measures – especially group involvement – are of interest to sociologists, as they may be related to civic engagement in general and political participation more specifically.

Table 21 contains descriptive statistics for measures related to marriage and family formation. There has been debate between sociologists and behavior geneticists about the heritability of sexual orientation (Bearman and Bruckner, 2002). Fertility has also been the subject of recent work debating the influence of nature and nurture (Rodgers et al., 2008).

Table 22 contains descriptive statistics for measures of socioeconomic class.

Table 23 contains descriptive statistics for measures of socio-political attitudes. Recently, several political scientists argued and found evidence for the idea that socio-political attitudes arrayed along the well-known dimension of liberal-to-conservative were partially rooted in biological differences between people (Alford et al., 2005).

Table 20: Descriptive statistics for measures of Sociability

Outcome	Coding	Mean	StDev	Min	Max
Positive relations with others	Sum of three items [a]	16.382	4.067	3	21
Number of group meetings attended in typical month	Sum of 5 items, each asking about the frequency of attending a different type of group meeting	5.508	7.127	0	60

Table 21: Descriptive statistics for measures of Marriage and Family

Outcome	Coding	Mean	StDev	Min	Max
Homosexual orientation	Based on 1 item, bisexuals and homosexuals combined since there were very few	0.032	0.176	0	1
Number of biological children	1 item	1.922	1.415	0	5
Marital conflict	Mean of 5 items, labeled marital risk scale by MIDUS investigators[a]	1.907	0.642	1	4
Respondent ever divorced	Based on answers to 3 items, equals 1 if ever divorced and 0 if not.	0.265	0.442	0	1

Table 22: Descriptive statistics for measures of Social Class

Outcome	Coding	Mean	StDev	Min	Max
Education	1 item	2.728	0.982	1	4
Socioeconomic index	Socioeconomic index of current job or last job if respondent not currently working [a]	37.425	14.175	9.56	80.5
Income	Mean of personal income response category range [a]	2.5e+04	2.7e+04	0	2.0e+05
Net worth	Mean of assets response category range	1.1e+05	2.1e+05	-1.7e+05	1.1e+06
Work complexity	Sum of 3 items	8.654	2.191	1	13
Autonomy	Sum of 4 items	11.692	3.614	1	17

Table 23: Descriptive statistics for measures of Socio-political Attitudes

Outcome	Coding	Mean	StDev	Min	Max
Social conservatism	Sum of 9 items	25.273	9.514	9	57
Economic libertarianism	Sum of 2 items	8.435	5.040	0	20
Attitudes about race	Sum of 4 items	6.082	3.063	1	13
Alienation	Sum of 6 items	21.586	6.456	6	42

Table 24: Descriptive statistics for measures of Social Obligations

Outcome	**Coding**	**Mean**	**StDev**	**Min**	**Max**
Felt obligation to give	Sum of 2 items	12.117	4.755	0	20
Dedication to family and friends	Mean of 8 items, labeled as normative primary obligation scale by MIDUS investigators[a]	59.757	13.064	0	80
Civic obligation	Mean of 4 items, labeled as normative civic obligation scale by MIDUS investigators[a]	30.498	8.155	0	40

Table 24 contains descriptive statistics for two measures of altruism.

Table 25 contains descriptive statistics for three measures of religion. The heritability of religiosity has been an active area of research, including a rare instance of behavior geneticists publishing in a sociology journal (Eaves et al., 2008).

3.1.2 Measures of environmental similarity

Tables 26 and 27 below contain the components of seven measures of environmental similarity. The seven measures are: similarity of childhood environment, proportion of life lived together, frequency of contact, frequency of virtual contact (e.g. mail, phone), psychological intimacy, negativity in the twin relationship, and advice seeking and giving. Proportion of life lived together was calculated as the ratio of the number of years lived with the twin to the age of the twin measured in years. Measures of contact, both real and virtual, were calculated by combining several nested questions about the frequency of contact between twins. The other measures of environmental similarity were calculated by simply taking the mean of all component measures. Following the construction of these measures, each was standardized in order to facilitate comparisons in the forthcoming analysis.

Table 25: Descriptive statistics for measures of religion

Outcome	**Coding**	**Mean**	**StDev**	**Min**	**Max**
Religiosity	Sum of 9 items, validated by factor analysis	18.981	6.995	1	29.1
Sectarianism	Sum of 3 items	4.791	2.453	1	9
Bible views	1 item, only asked of those who identified with a Christian religion	3.577	1.406	1	5

I measure environmental similarity in a more comprehensive way than other published analysis. I incorporate measures of environmental similarity that are influenced most by the twins themselves, as well as measures of the environment that are predominantly influenced by others. I also include measures of the quality of the twin relationship, as well as the quantity of time twins spend together. Previous research testing the equal environments assumption focused mainly on the quantity of time twins spent together. But twins who see each other rarely may remain emotionally close and may exert substantial influence on each other. Co-twins who are physically distant may still remain central members of each other's reference group.

One additional aspect of childhood environmental similarity, the level of equality perceived by the twins, was included in the analysis. Each twin was asked, When you were children, and just the two of you were doing things together, which one of you was more likely to be the leader and which one the follower? Twins were given the option of choosing a third option by indicating that the relationship was about equal. I coded twin responses to this question into the following set of categories that exhausts the possible combinations: (1) twins agree about inequality in the relationship; (2) twins agree about equality in the relationship; (3) both twins report they tended to be the leader; (4) both twins report they tended to be the follower; (5) one twin reported equality, and the other reported being a leader, and finally, (6) one twin reported equality, and the other reported being a follower. In some cases, in order to simplify the analysis, the measure of inequality in the relationship was collapsed into three categories (1) agreed about equality, (2) agreed about inequality and (3) other.

While the vast majority of questions were asked of all twins, there are some

Table 26: Components of scales measuring environmental similarity

Scale	**Question**	**Mean**	**StDev**	**Min**	**Max**
Similarity of childhood	When you were children, how often did you and your twin have the same playmates?	1.766	0.789	1	4
	How often were you in the same classroom in school?	2.296	1.091	1	4
	How often did you dress alike?	2.283	0.994	1	4
	Did you ever share the same bedroom with your twin during the time you were growing up?	0.981	0.137	0	1
	For how many years did the two of you have the same bedroom?	16.299	4.659	1	70
Proportion of life lived together	Are you currently living with your twin?	0.033	0.178	0	1
	In your lifetime, how many total years have the two of you lived together?	19.094	3.286	0	50
	Age of twins	44.683	12.135	25	74
Frequency of contact	How frequently do you and your twin see each other? (combined from several connected questions)	3.799	2.243	1	9
Frequency of virtual contact	Not including visits, how frequently are you in contact with your twin – including phone calls, letters, or electronic mail messages? (combined from several connected questions); only asked of twins who did not live together or see each other	4.726	1.943	1	9

Table 27: Components of scales measuring environmental similarity, continued

Scale	**Question**	**Mean**	**StDev**	**Min**	**Max**
Psychological Intimacy	How much does your twin understand the way you feel about things?	1.392	0.734	1	4
	How much can you rely on (him/her) for help if you have a serious problem?	1.253	0.644	1	4
	How much does your twin really care about you?	1.092	0.366	1	4
	How much can you open up to your twin if you need to talk about worries?	1.357	0.730	1	4
Negativity in twin relationship	How much does your twin criticize you?	2.933	1.006	1	4
	How often does your twin make too many demands on you?	3.283	0.812	1	4
	How often does your twin let you down when you are counting on (him/her)?	3.461	0.760	1	4
	How often does your twin get on your nerves?	2.958	0.938	1	4
Advice seeking and giving	When you have a personal or practical problem, how much of the time do you turn to your twin for advice or help?	3.142	1.284	1	5
	When your twin has a personal or practical problem, how often does (he/she) turn to you for advice or help?	3.167	1.251	1	5

questions that were asked of only a sub-sample. Questions about virtual contact were not asked of twins who said they lived together, or of twins who reported seeing each other on at least a daily basis.

3.1.3 Measure of zygosity

My measure of zygosity was provided by MIDUS investigators. As in most twin surveys, MIDUS investigators constructed their measure of zygosity based on questions asked of the twins about their level of physical similarity and about how often people confused them when they were growing up. While a biological assay is the gold standard for determining zygosity, the validity of zygosity measures based on survey questions is high. Validation studies have estimated the accuracy of survey-based methods at over 90% (Kasriel and Eaves, 1976). 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. The other 1.8% whose zygosity was coded as undetermined by MIDUS investigators, are not considered here.

3.1.4 Other control variables

Following the convention in twin studies, I also control for age and gender in my analyses. It was important to control for gender here, since, by chance, gender is related to zygosity in these data. DZ twins were significantly more likely than MZ twins to be female.

3.2 Handling missing data

Cases missing on one of the measures of environmental similarity were excluded from all analyses. Missingness on outcome variables was a different story. Listwise deletion of all cases that were missing on any of the fifty-eight outcomes [11] would have reduced the sample by about 64%. Excluding several variables with large percentages of missing data, listwise deletion would still reduce sample size by about 44%. Thus, I decided not to listwise delete cases on outcome variables.

Instead, all cases that had data available on any particular outcome measure were included in the analysis for that outcome. Hence, the forthcoming comparison of analyses across dependent variables is also by necessity a comparison of analyses of somewhat different sub-samples of the data. It is possible that some of the differences

[11] Listwise deletion would not include the fifty-ninth outcome variable, views of the bible, since this question was only asked of people who identified as Christian

in the results for different variables stem from chance differences in the samples. However, the overlap between the samples is very substantial. In addition, I will draw conclusions only from strong patterns in the data, thereby limiting the likelihood that the summary of the results will be affected by chance differences between subsamples.

4 Methods

4.1 Comparisons of means and variances by zygosity

It is customary in analyses of twins to ensure that variables are measured similarly across categories of zygosity. Thus, the first step in the analysis is to compare means and variances of each outcome variable between MZ twins and DZ twins. I compare means with t-tests and I compare variances using Levene's test (Levene, 1960).

4.2 Examining measures of environmental similarity

The second step in the analysis is to examine the measures of environmental similarity. As with any variable in twin analysis, we have reports of environmental similarity from each twin. Simply taking the average of twin reports seems to be a natural choice. First, however, in order to examine the reliability of these measures, I examined the correlation between twin reports on each measure of environmental similarity. I then did this separately by zygosity. MZ twins might have more similar recollections of the extent to which they shared their environment.

4.3 Absolute value regressions

After examining the key independent variables, I turn to the main task of examining the extent to which environmental similarity confounds the effects of zygosity. I approach this problem in a number of ways. First, I adopt a method used by sociologists (Horwitz et al., 2003) to test the twin methodology. I take the absolute value of the difference on each trait for each twin pair, and regress this value on a dummy variable for zygosity. Then I run a similar regression, now controlling for one of the measures of environmental similarity. The absolute value regressions are shown in the equations below.

$$|Y_1 - Y_2| = b_{10} + b_{11}(mz) + b_{12}(age) + b_{13}(male) + e_1 \quad (4)$$

$$|Y_1 - Y_2| = b_{20} + b_{21}(mz) + b_{22}(age) + b_{23}(male) + b_{24}(environment) + e_2 \quad (5)$$

I compare the coefficient for zygosity in the first regression that does not include a control variable for environmental similarity (b_{11}) with the coefficient for zygosity in the second regression (b_{21}) that includes one of the eight controls for environmental similarity. I calculate the percentage change in the coefficient from the first equation $(b_{11}-b_{21})/b_{11}$. I test the significance of the difference between the zygosity coefficients b_{11} and b_{21} with a test described in Clogg et al. (1995).

These sets of regressions are conducted for each outcome variable. The functional forms of the regressions are chosen in accordance with the distribution of the outcome variable. For dichotomous outcome measures, logistic regression was used. For ordinal outcomes with five or fewer categories, I used ordinal logistic regression. For integer outcomes with more than five categories that were heavily skewed to the right, I ran either a Poisson or negative binomial regression. Poisson regression was chosen if the variance of the outcome was equal to the mean; otherwise negative binomial regression was run. Finally, non-integer outcomes with many categories were logged and run with linear regression.

4.4 Defries-Fulker regressions

As another window onto the confounding effects of environmental similarity, I employed a method that has been popular among behavior geneticists, the DeFries-Fulker (DF) regression (DeFries and Fulker, 1985). This method is not appropriate for use with dichotomous outcomes. For dichotomous outcomes, I used a DF-like method developed for binary outcomes by Kohler and Rodgers (1999). First, I describe the canonical DF model, and then I describe the DF-like method for binary outcomes.

DF models produce estimates of h^2 and c^2 as regression coefficients. First, a variable representing heritability called h is created, which is equal to .5 for DZ twins and 1 for MZ twins. Then, the trait for twin one is regressed on three variables: h, the trait for twin two, and the interaction between h and the trait for twin two. The coefficient on the interaction term is an estimate of heritability (h_2) and the coefficient for twin two's trait is an estimate of c^2. The equation for a simple DF regression is:

$$Y_1 = b_0 + b_1 Y_2 + b_2 h + b_3 Y_2 h + e \tag{6}$$

As stated above, b_3 is an estimate of heritability. In order to control for environmental similarity, I add an interaction term to the DF regressions between twin two's trait and one of the measures of environmental similarity. The equation becomes:

$$Y_1 = b_0 + b_1 Y_2 + b_2 h + b_3 Y_2 h + b_4 X_{environment} + b_5 Y_2 X_{environment} + e \qquad (7)$$

Equation 7 includes an interaction term between a measure of environmental similarity ($X_{environment}$) and a trait for twin two (Y_2). A comparison of b_3 in equation 6 with b_3 in equation 7 measures the extent to which environmental similarity confounds the apparent effect of genes.

Speaking in terms of statistical interactions, I am comparing the strength of two different moderators of the relationship between the traits of twin one and two: the h variable (equal to .5 for DZ twins and equal to 1 for MZ twins) reflecting the effect of genes, and the measure of environmental similarity. As with the regressions of the absolute value of the difference, I examine the change in heritability estimates with the inclusion of controls for environmental similarity. Again I look at the percentage change in the coefficients, and test the significance of the difference with a Clogg test.

It should be noted that the ordering of the twins is arbitrary, and hence arbitrary changes in the estimates of heritability would result from changes in the ordering of some twin pairs. The customary solution for this problem is to double enter the data by making a copy of data from twins one and two, reversing the order of the twins, and then appending the resulting reversed copy to the original data. I adopt this solution here. Because the data are entered twice, I use the cluster() option in Stata to adjust for non-independent observations, as suggested by Kohler and Rodgers (2000a). Kohler and Rodgers (2000a) recommended the use of Generalized estimating equations (GEE) to achieve maximum efficiency of estimates. However, their simulations showed that the reductions in standard errors achieved by their methodology were very small, by their own admission. Hence, the benefits of GEE models seemed minute, and regressions were estimated with standard tools, albeit with adjustments for the non-independent data.

As I mentioned earlier, the DF method is not appropriate for dichotomous outcomes. For binary outcomes, I implemented a bivariate probit model with a Stata module written by Kohler and Rodgers (2000b) and described further in (Kohler and Rodgers, 1999). In the bivariate probit model, heritability estimates of dichotomous outcomes are not estimated directly. Instead, we assume that each dichotomous outcome reflects an underlying, unmeasured (latent) continuous variable. For example, one of the dichotomous outcomes, depression diagnosis, is assumed to reflect an underlying continuous distribution of well-being. It is assumed that a diagnosis of depression is assigned to people who score below a certain (estimated) threshold on the continuous latent measure of well-being. The continuous distribution underlying each dichotomous variable is assumed to be normally distributed. The normality assumption is made for computational convenience.

In this bivariate probit model, at least four parameters are estimated: the value for the threshold on the continuous variable that corresponds to a change in the dichotomous outcome, the correlation between the two variables underlying the dichotomous outcomes for each twin, an estimate of heritability, and an estimate of the influence of shared environment.

The bivariate probit version of the DF model is estimated by maximizing the following likelihood equation:

$$\log L = \Sigma_{i=1}^{n} \log \Phi_2 \left(q_{1i} \left(x_i \beta + \tau_0 \right), q_{2i} \left(x_i \beta + \tau_0 \right), q_{1i} q_{2i} \rho \right) \quad (8)$$

where Φ_2 is the cumulative bivariate normal distribution.

For monozygotic twins, $\rho = h^2 + c^2$. For dizygotic twins, $\rho = .5h^2 + c^2$. The values of all control variables included in the equation are included in x_i. Control variables include age and gender in addition to the variables reflecting environmental similarity described above.

The variables q_{1i} and q_{2i} equal +1 or -1, and hence determine the sign of each of the three parameters for the bivariate normal distribution. $q_{1i} = 1$ when the value on the dichotomous outcome for the first twin is equal to 1. $q_{1i} = -1$ when the value on the dichotomous outcome for the first twin is equal to 0. Similarly, $q_{2i} = 1$ when the value on the dichotomous outcome for the second twin is equal to 1, and $q_{2i} = -1$ when the value on the dichotomous outcome for the second twin is equal to 0. Note the third parameter of the bivariate normal distribution: $q_{1i} q_{2i} \rho$. When twins are concordant on the dichotomous outcome, then either both $q_{1i} = 1$ and $q_{2i} = 1$ or $q_{1i} = -1$ and $q_{2i} = -1$. Hence, the contribution of a twin pair to the third parameter will be positive when the twins are concordant on the outcome variable. When twins are discordant on the dichotomous outcome variable, then their contribution to the third parameter will be negative.

The program used to estimate DF-like regressions for binary outcomes, Kohler and Rodgers (2000a), uses Stata's ml command (Gould et al., 2003).

4.5 Propensity score matching

In the previous sets of regressions, I examined the effect of each measure of environmental similarity in isolation from the others. In a final examination of the confounding effect of environmental similarity, I compared the results of standard DF regressions with the results of DF regressions run on data matched on all measures of environmental similarity. Literal matching on all environmental similarity measures would be impossible. Given the number of combinations of values on different environmental variables, no pair of twins shared exactly the same values with any other pair.

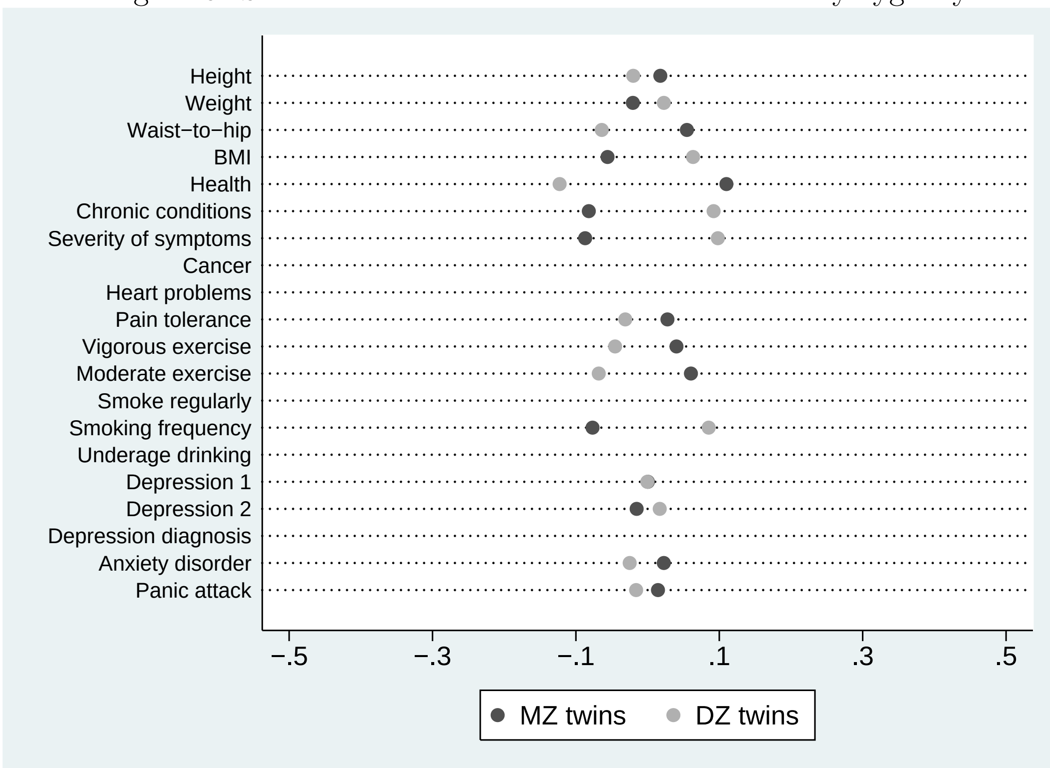
The solution here is propensity scores (Morgan and Winship, 2007). In order to match MZ and DZ twins on levels of environmental similarity, I used the psmatch2 command in Stata. This command runs a probit regression of zygosity on all environmental similarity measures. Based on the results of the probit regression, a probability of being monozygotic twins is estimated for each pair. This estimated probability is the propensity score that psmatch2 uses to develop weights to match the DZ twins to the MZ twins. The data are weighted such that the differences in environmental similarity between MZ and DZ twins disappear. Essentially, DZ twins who share more similar environments are weighted up, while DZ twins whose environments are dissimilar are weighted down.

Propensity score matching has several advantages over controlling by regression. First, the process of controlling for variables is made much more transparent by matching than by regression. Through the graphs that I will show, the extent of differences in environmental similarity by zygosity becomes very clear. In addition, in this context, the choice matching rather than controlling by regression increases power. The alternative would be to add at least ten interaction terms to the DF regression, one for each of the seven ordered environmental variables, and several for the measure of equality.

5 Results

The results of the analysis are organized into five sections. In the first part of the Results section, I examine whether the means and variances of the continuous outcome measures differ by zygosity. In the second part of this section, I examine whether means on measures of environmental similarity differ by zygosity. The third part of the Results section shows how much heritability estimates are reduced by controlling for each measure of environmental similarity at a time in regressions predicting absolute co-twin differences in the outcomes. The fourth part of the Results section shows the extent to which heritability estimates are reduced by controlling for each measure of environmental similarity at a time in Defries-Fulker regressions. In the last section, I show the extent to which estimates of heritability change when MZ and DZ twins are matched on all environmental measures.

Figure 8: Standardized means of outcome variables by zygosity



5.1 Comparisons of means and variances by zygosity

Figures 8, 9, and 10 present comparisons of means by zygosity for the 52 examined outcomes that were not dichotomous. Variables were standardized so that they could be displayed on a common metric. The units of measurement are standard deviations. The figures show that the average differences between MZ and DZ twins on these traits rarely exceed one fifth of a standard deviation, and are often considerably smaller than that.

Differences on most traits were not significant, with the following notable exceptions: health (self-reported), chronic conditions, severity of symptoms, moderate exercise, smoking frequency, mental health (self-reported), years of education, and net worth.

MZ twins are better off compared to DZ twins in all of the ways for which differences are significant. MZ twins tend to be healthier, exercise more, smoke less, report better mental health, more years of education, and a higher net worth, than DZ twins. But these differences are relatively small.

Figure 9: Standardized means of outcome variables by zygosity, cont.

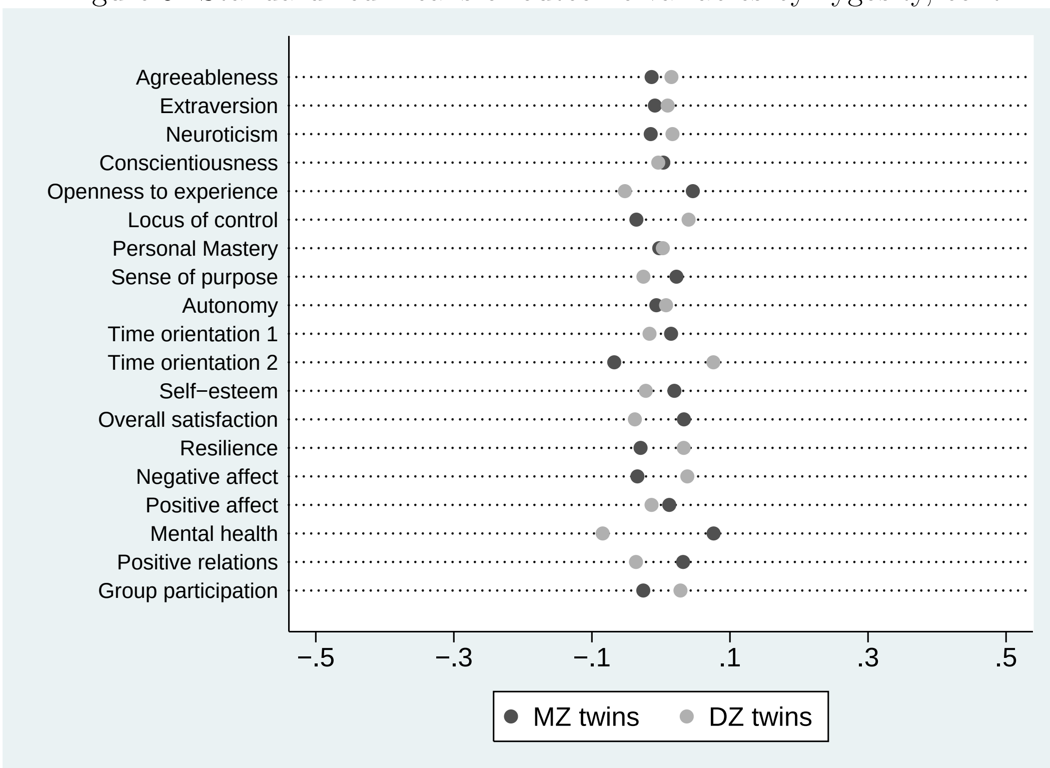


Figure 10: Standardized means of outcome variables by zygosity, cont.

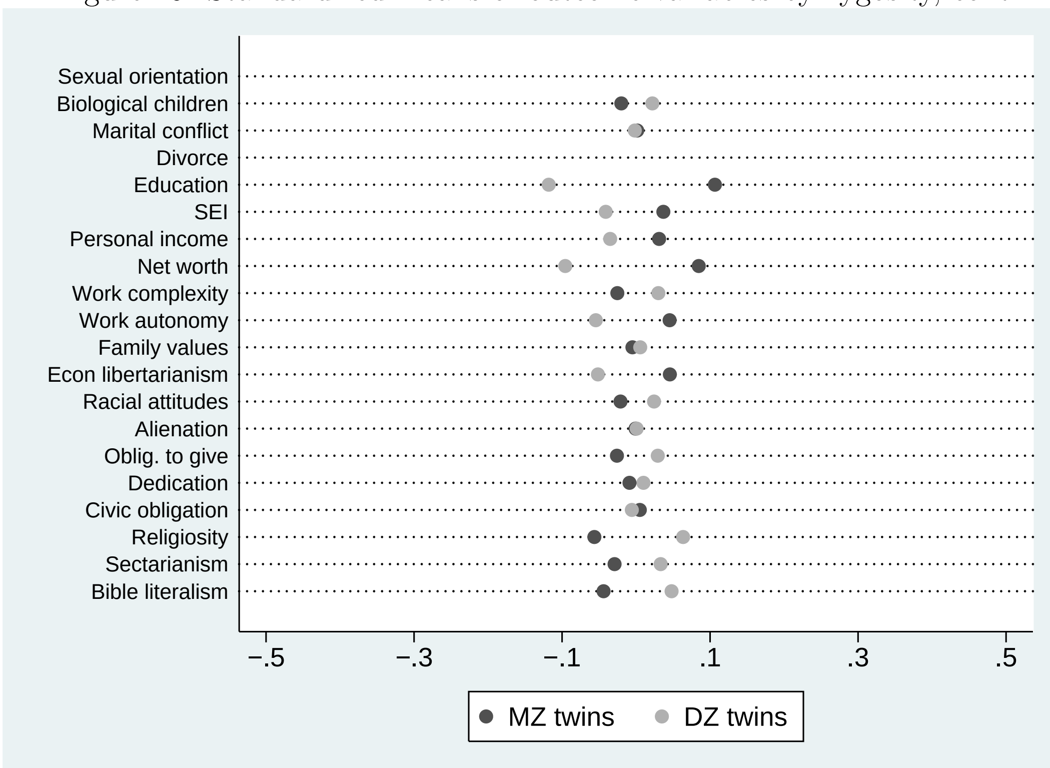


Table 28: Standardized outcome variances by zygosity

Outcome	**MZ Var**	**DZ Var**	**p-value**
Height	0.955	1.051	0.750
Weight	0.970	1.034	0.806
Waist-to-hip	1.008	0.985	0.664
BMI	0.882	1.129	0.015
Health	0.898	1.087	0.457
Chronic conditions	0.862	1.140	0.391
Severity of symptoms	0.917	1.077	0.031
Pain tolerance	0.953	1.053	0.298
Vigorous exercise	1.000	0.998	0.791
Moderate exercise	0.959	1.039	0.068
Smoking frequency	0.839	1.166	0.022
Depression 1	1.016	0.984	0.801
Depression 2	0.875	1.139	0.006
Anxiety disorder	1.231	0.745	0.007
Panic attack	1.058	0.937	0.003
Agreeableness	0.983	1.021	0.688
Extraversion	1.013	0.987	0.628
Neuroticism	1.055	0.939	0.593
Conscientiousness	0.980	1.024	0.283
Openness to experience	0.954	1.048	0.159

In Tables 28 and 29, I compare variances by zygosity for each of the 52 non-dichotomous outcomes. Again, variables were standardized so that one can gauge whether and how differences in the variances of MZ and DZ twins vary by outcome. Levene's test for differences in variances was performed, and the resulting p-values are displayed in the right-most column for each outcome.

When variances are significantly greater for MZ twins than for DZ twins, the equal environment assumption might be violated. MZ twins may experience a more similar shared environment that pulls both of them farther away from the average on some trait. This phenomenon could in turn result in a higher variance for MZ twins than for DZ twins.

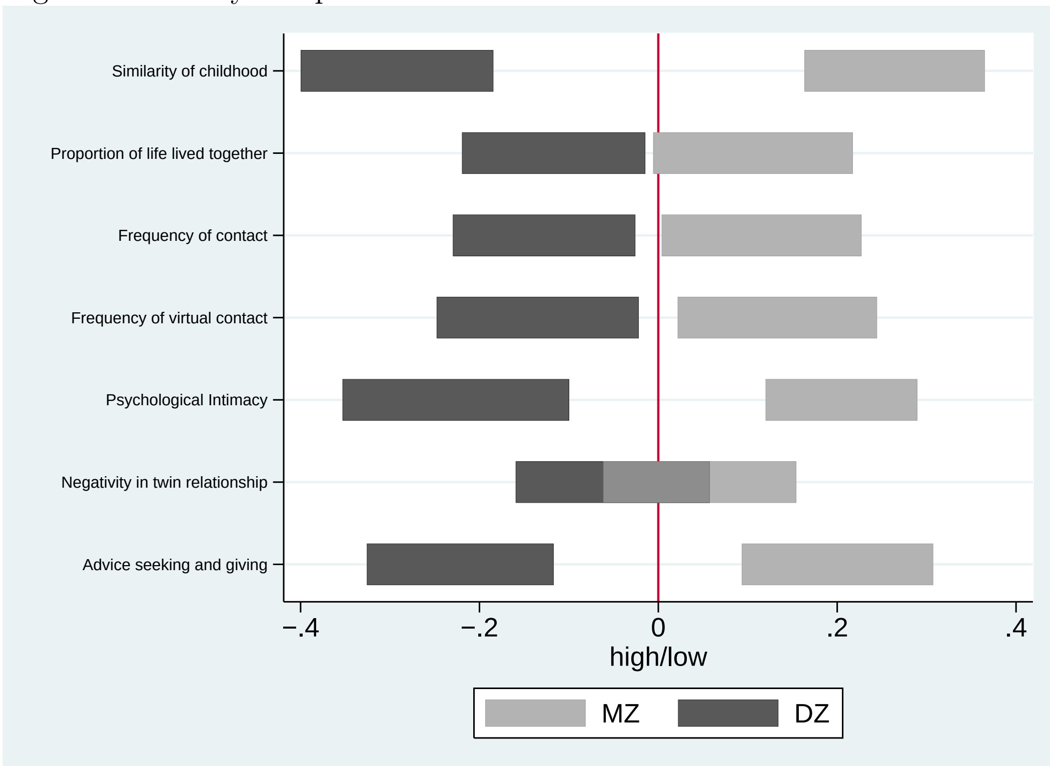
There is little evidence for this conjecture here, however. Differences in the variances of the outcomes by zygosity are rarely significant.

And when there are significant differences in variances by zygosity, the ordering of the differences is often not what one would expect if the equal environments assumption were violated.

Table 29: Standardized outcome variances by zygosity, continued

Outcome	**MZ Var**	**DZ Var**	**p-value**
Locus of control	0.953	1.051	0.480
Personal Mastery	0.994	1.009	0.488
Sense of purpose	0.981	1.023	0.508
Autonomy	1.018	0.981	0.267
Time orientation 1	0.968	1.037	0.423
Time orientation 2	0.897	1.106	0.019
Self-esteem	1.077	0.914	0.996
Overall satisfaction	0.996	1.003	0.435
Resilience	1.049	0.945	0.572
Negative affect	1.003	0.996	0.971
Positive affect	0.989	1.014	0.969
Mental health	0.954	1.039	0.011
Positive relations	0.966	1.037	0.083
Group participation	0.858	1.157	0.006
Biological children	1.018	0.981	0.286
Marital conflict	1.031	0.968	0.583
Education	0.910	1.075	0.002
SEI	0.940	1.065	0.418
Personal income	0.920	1.090	0.937
Net worth	1.233	0.721	0.000
Work complexity	1.003	0.997	0.741
Work autonomy	0.934	1.076	0.179
Family values	0.994	1.008	0.066
Econ libertarianism	1.021	0.973	0.565
Racial attitudes	1.028	0.968	0.458
Alienation	1.058	0.935	0.660
Oblig. to give	1.015	0.983	0.282
Dedication	1.037	0.960	0.057
Civic obligation	0.981	1.023	0.290
Religiosity	1.075	0.909	0.255
Sectarianism	1.036	0.959	0.625
Bible literalism	1.031	0.964	0.476

Figure 11: Ninety-five percent confidence intervals of environmental similarity



5.2 Examining measures of environmental similarity

Now, I turn to the measures of environmental similarity. Recall that there were eight of these, including seven quantitative and one qualitative. In Figure 11, I present 95% confidence intervals for the means of the seven scales of environmental similarity. I do this separately for MZ and DZ twins. The scales were standardized to facilitate interpretation. MZ twins report greater similarity on six out of the seven scales. With regard to the scale measuring negativity in the twin relationship, the mean for MZ twins is higher than the mean for DZ twins, but the difference between these means is not significant. Zygosity appears to have the largest effect on similarity of childhood. MZ twins report a much more similar childhood environment than do DZ twins.

Advice giving and receiving is also more common between MZ co-twins than between DZ co-twins. MZ twins report higher levels of psychological intimacy with their co-twin than do DZ twins. Zygosity differences in frequency of contact and in the proportion of life lived together, are significant but less pronounced.

I now examine the extent of agreement between twin reports of their level of environmental similarity. For example, when twin one reported that he or she spent a lot of time with his or her co-twin, did the co-twin also report spending a lot of time together? Figure 12 displays correlations between co-twin reports on these characteristics, separately by zygosity. If twin reports on their environment are highly divergent, then the validity of these reports is in doubt. The level of agreement between twins is a rough gauge of the reliability of the measures of environmental similarity, and as such, also an upper bound on the validity of these measures (Alwin, 2007).

Twins reports on the proportion of life lived together, and on the level of current contact are largely consistent with each other. Co-twins agree to a much lesser extent about the level of psychological closeness that characterizes the relationship and in terms of perceptions of negativity in the relationship. In general, twins agree more about objective, current aspects of the environment they share with their twin. Twins agree less about events occurring a long time in the past (e.g. childhood environment) and about psychological components of the twin relationship.

MZ twins were no more likely than DZ twins to agree about the amount of current contact, or about the proportion of life lived together. But there was considerably greater accord between MZ than DZ twins with regard to psychological aspects of the relationship. This is unsurprising, given that research has found that MZ twins to be psychologically closer to their co-twins than are DZ twins.

Figure 13 presents data about twin perceptions of their status in the relationship when growing up. Interestingly, a majority of twins perceived some sort of status inequality in the relationship growing up. There was also a surprising amount of disagreement about who tended to lead and who tended to follow. Twins were in agreement only about 60% of the time.

As shown in Figure 14, disagreement about relative status growing up was only slightly higher between DZ co-twins than between MZ co-twin pairs. The zygosity differences in perceptions of social status are especially modest, given how much more MZ twins agree about level of closeness and level of negativity than do DZ twins.

Figure 12: Agreement between twins on measures of environmental similarity

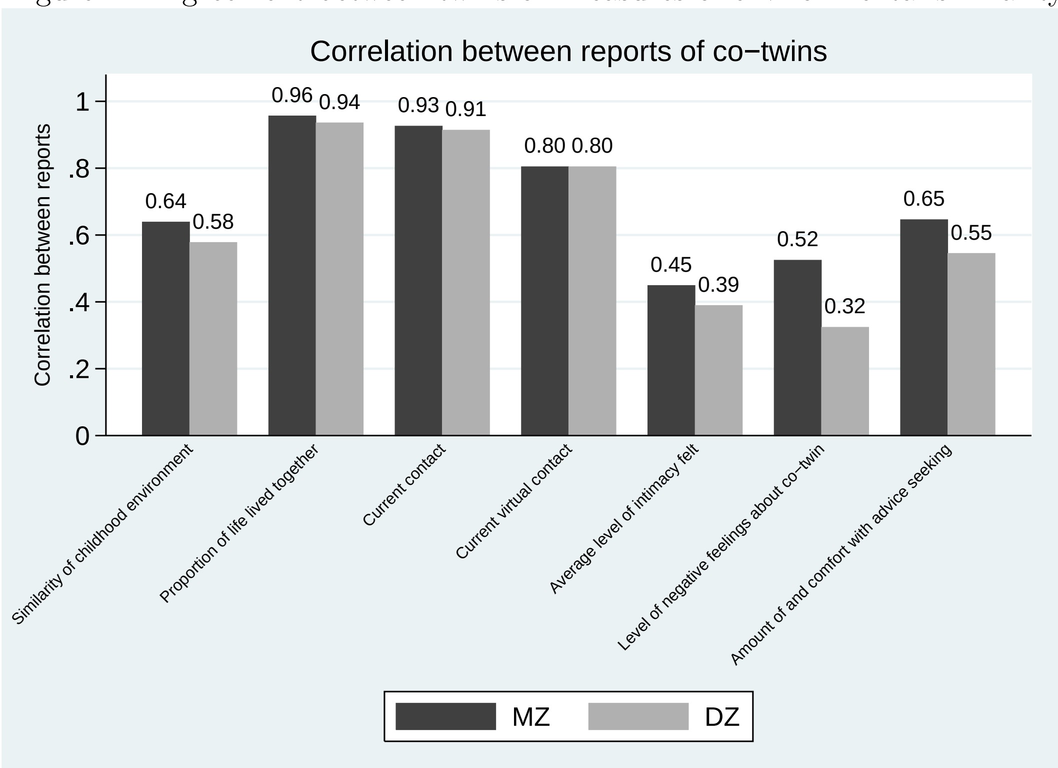


Figure 13: Twin reports about equality in relationship growing up

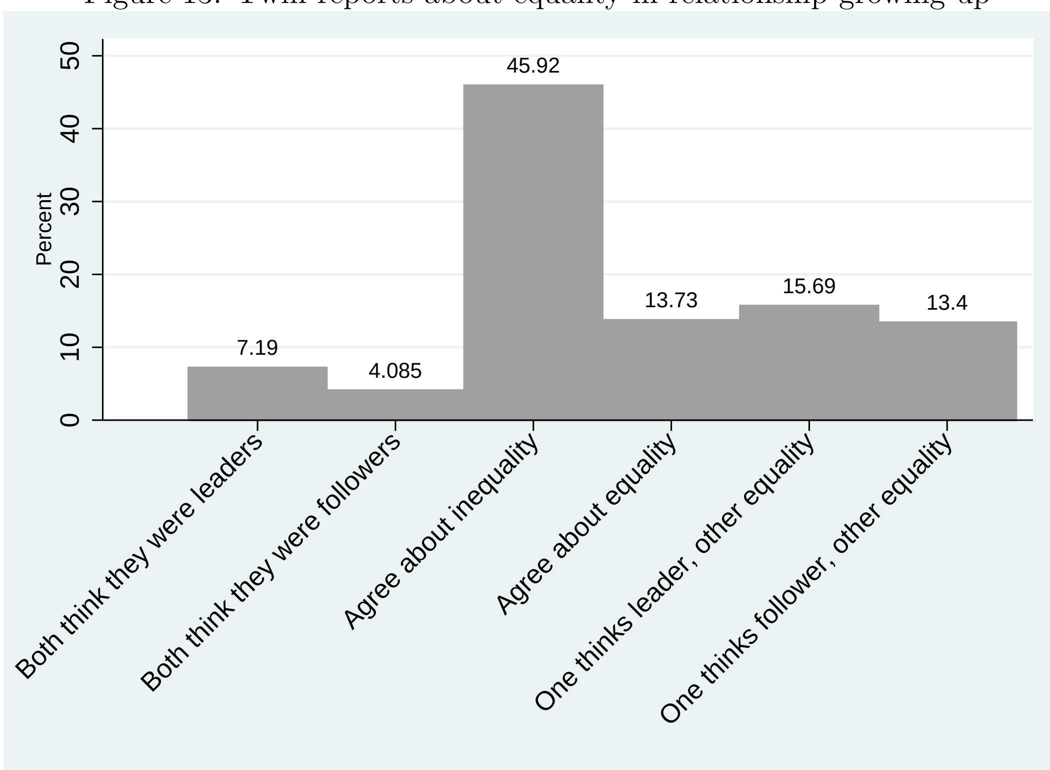
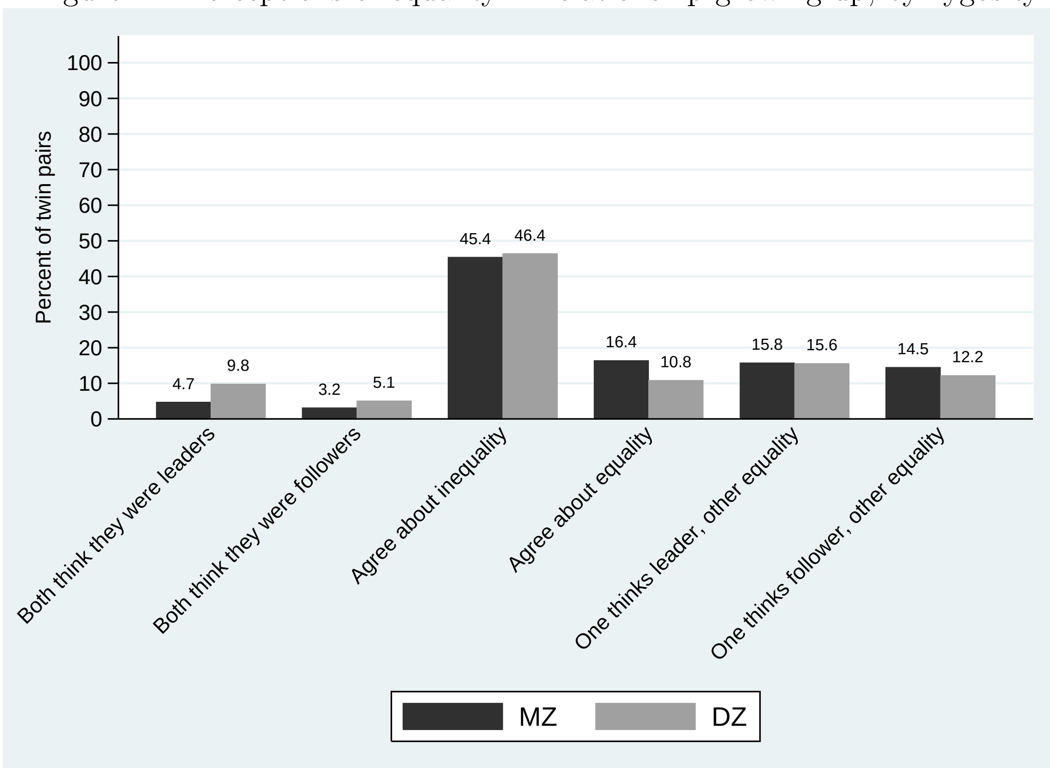


Figure 14: Perceptions of equality in relationship growing up, by zygosity



5.3 Absolute value regressions

Having examined measures of environmental similarity by themselves, I now evaluate the impact of these measures as control variables in regressions. I examined the effects of each of the eight control variables on heritability estimates. I ran regressions of the absolute difference between co-twins in 59 outcomes with and without each control variable.

Figures 15-22 present the most important piece of information from these regressions: the percentage decrease in the coefficient for zygosity when measures of environmental similarity are controlled. When the change was significant at $p \leq .05$ according to the Clogg test, the bar is shaded black. Grey bars indicate changes that were not significant. Note that in some cases, the percentage change in the zygosity coefficient was positive, indicating that the apparent influence of genes actually increased when the measure of childhood environmental similarity was controlled. In order to conserve space I do not present bars for outcomes in cases where the coefficient for zygosity increased. Instead, these outcomes are listed in a caption at the bottom of each figure.

Figure 15 displays the percentage change in the coefficient for zygosity when controlling for a measure of childhood environment. As expected, the coefficient for zygosity usually decreased in size when the measure of childhood environmental similarity was included in the regression. However, the effect of partialing out childhood similarity was usually not significant. In addition, there does not appear to be any strong pattern across different types of outcomes. One might expect that physical characteristics and health would be more robust to the effects of childhood environment than traits like attitudes and social class. But the effect of controlling for childhood environmental similarity is not patterned in this way. Surprisingly, the genetic effect on health appears to be completely confounded with childhood environmental similarity. Much of the apparent effect of genes on sexual orientation and on socioeconomic status (SEI) as an adult appear to be spurious. Given the large number of outcomes examined, however, one is bound to find significant effects if one is looking for them.

Figure 16 displays results that are analogous to those presented in Figure 15, except that I control for the measure of perceived equality growing up. Equality in the twin relationship does not seem to confound the apparent effect of zygosity at all. This is not surprising, given that the relationship between zygosity and relationship equality was shown to be very weak.

Figure 17 displays the percent change in the coefficient for zygosity when controlling for the amount of current contact between twins. For most outcomes, the zygosity coefficient was reduced by less than 20% or not at all. However, as in Fig-

Figure 15: Percent change in coefficient for zygosity when controlling childhood environment

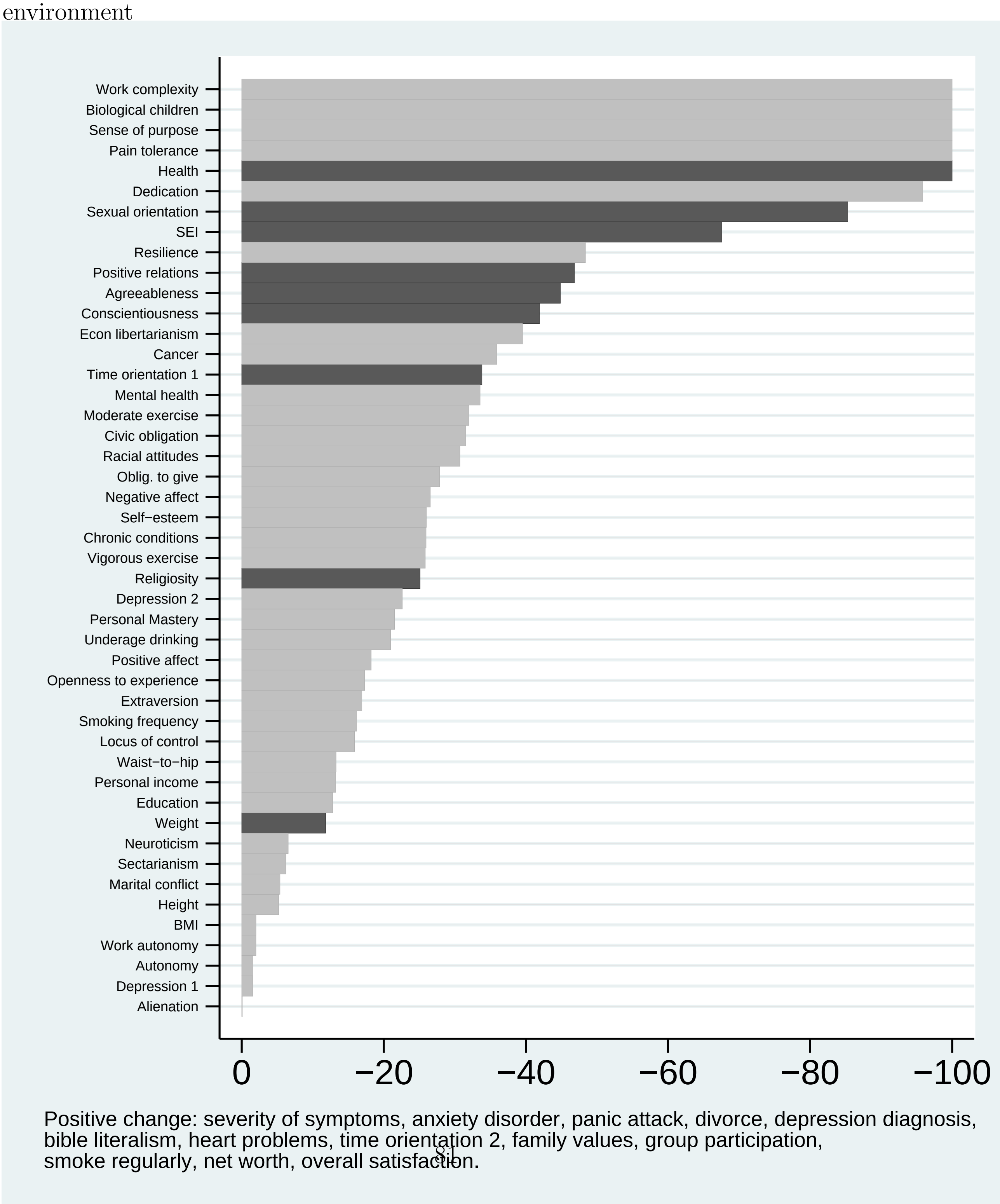


Figure 16: Percent change in the coefficient for zygosity when controlling for status inequality in relationship growing up

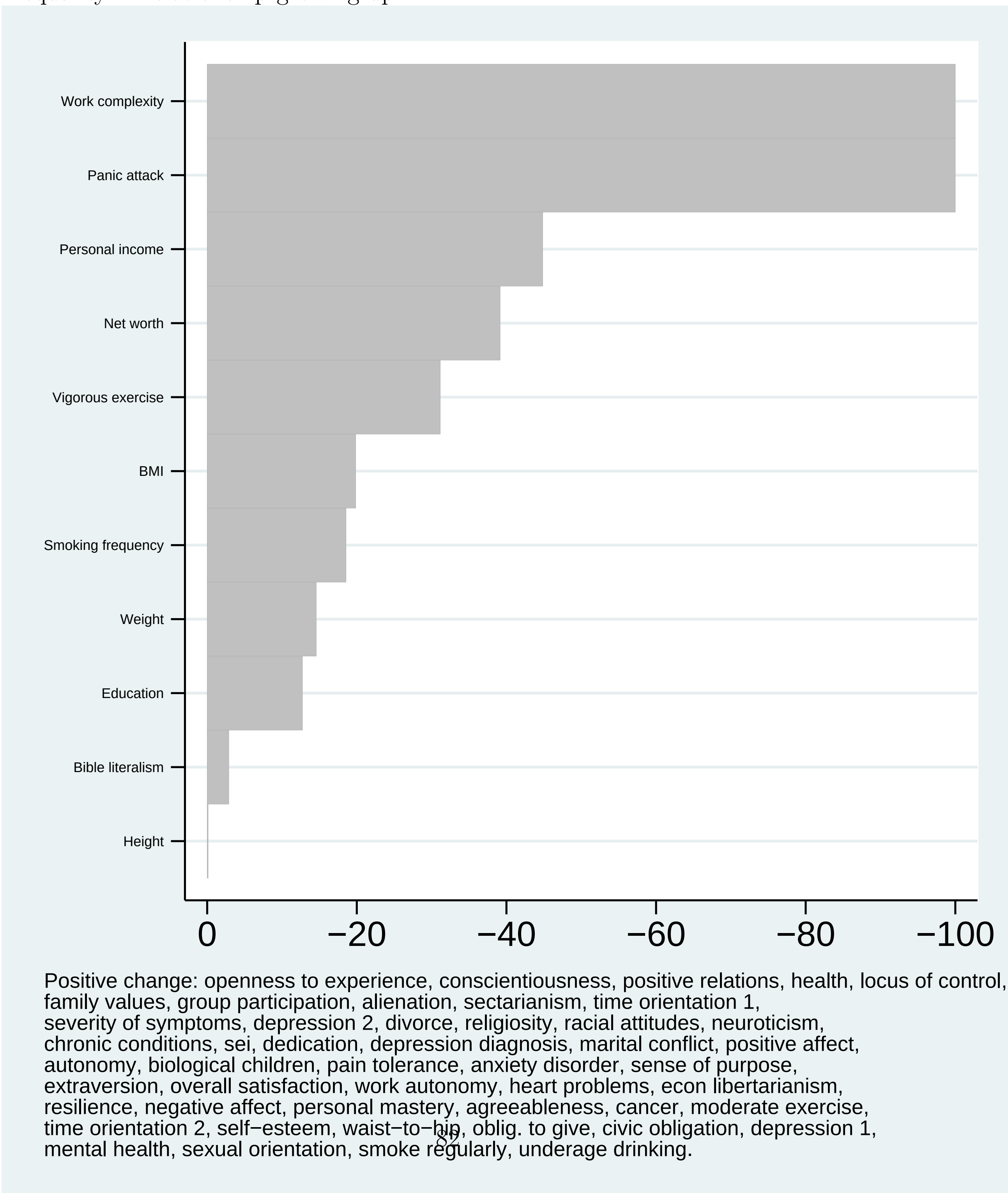
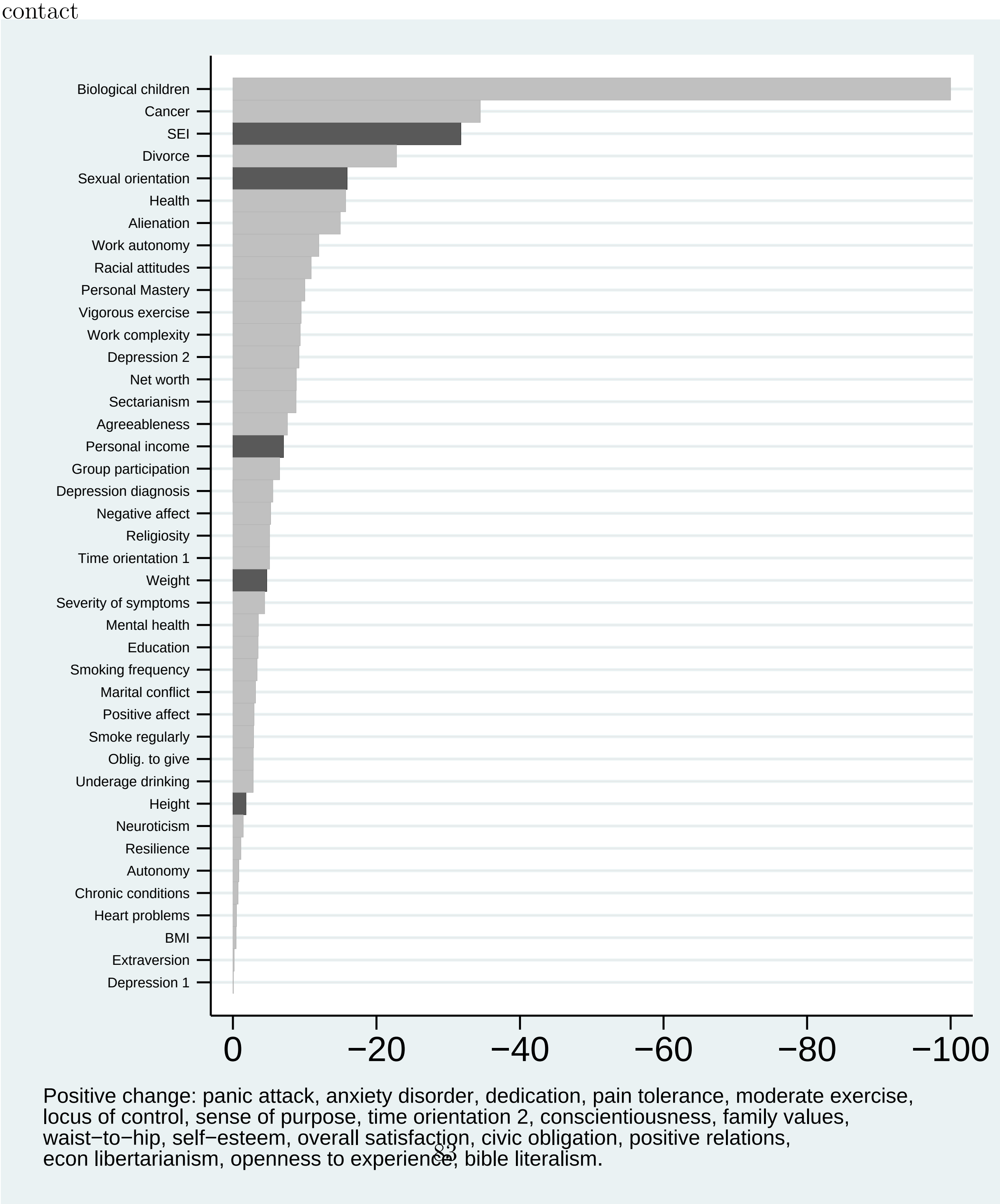


Figure 17: Percent change in the coefficient for zygosity when controlling for current contact



ure 15, the percentage reductions for the zygosity coefficient with respect to sexual orientation and the socioeconomic index (SEI) were both substantial and significant. Otherwise, there does not appear to be any strong pattern of results across categories of the dependent variables.

Figure 18 displays the percent change in the coefficient for zygosity when controlling for the proportion of the twins' lives spent living together. The effect of zygosity appears to be confounded with this measure of environmental similarity for a variety of unrelated outcomes: number of children, sense of purpose, the severity of symptoms of illness, the propensity to divorce, one measure of depression, the propensity to receive a diagnosis of depression, and self-reported health. The change in the coefficient for the measure of depression was significant at $p \leq .10$. There is preliminary evidence that the apparent effect of genes on depression is at least partially spurious. That said, the change in the zygosity coefficient was insignificant and often small in size for the majority of outcomes.

Figure 19 shows the percent change in the zygosity coefficient when a control for virtual contact (e.g. phone and mail) is included. Recall that these regressions were run on a smaller sample than the regressions testing the effects of other measures of environmental similarity, since questions about virtual contact were not asked of respondents who said they saw their twin on a daily basis. Given the reduction in sample size, perhaps it is not surprising that none of the changes in zygosity coefficients are significant in Figure 19.

Figure 20 shows the percent change in the coefficient for zygosity when a control for psychological intimacy is included in the regression models. The effect of genes appear to be confounded with the effects of psychological intimacy on traits related to do with psychological characteristics that are important for getting along with others. The apparent effect of genes on anxiety disorders virtually disappears when controlling for psychological intimacy. In addition, the apparent effect of genes on one measure of depression, as well as on the propensity to receive a diagnosis of depression, was substantially reduced by a control for psychological intimacy between twins. The reduction in the coefficient in the regression predicting the propensity for a depression diagnosis was significant at $p \leq .10$. Controlling for psychological intimacy also substantially reduces the apparent effects of genes on marital conflict and on positive relations with others. In general, psychological intimacy between twins seems to be confounded with the effects of genes on outcomes that have to do with the temperance of mind that is important for getting along with others.

Figure 21 shows the percent change in the coefficient for zygosity when a control for negativity in the twin relationship is included in the regression models. Negativity in the relationship does not appear to play a significant confounding role.

Figure 18: Percent change in the coefficient for zygosity when controlling for the proportion of life lived together

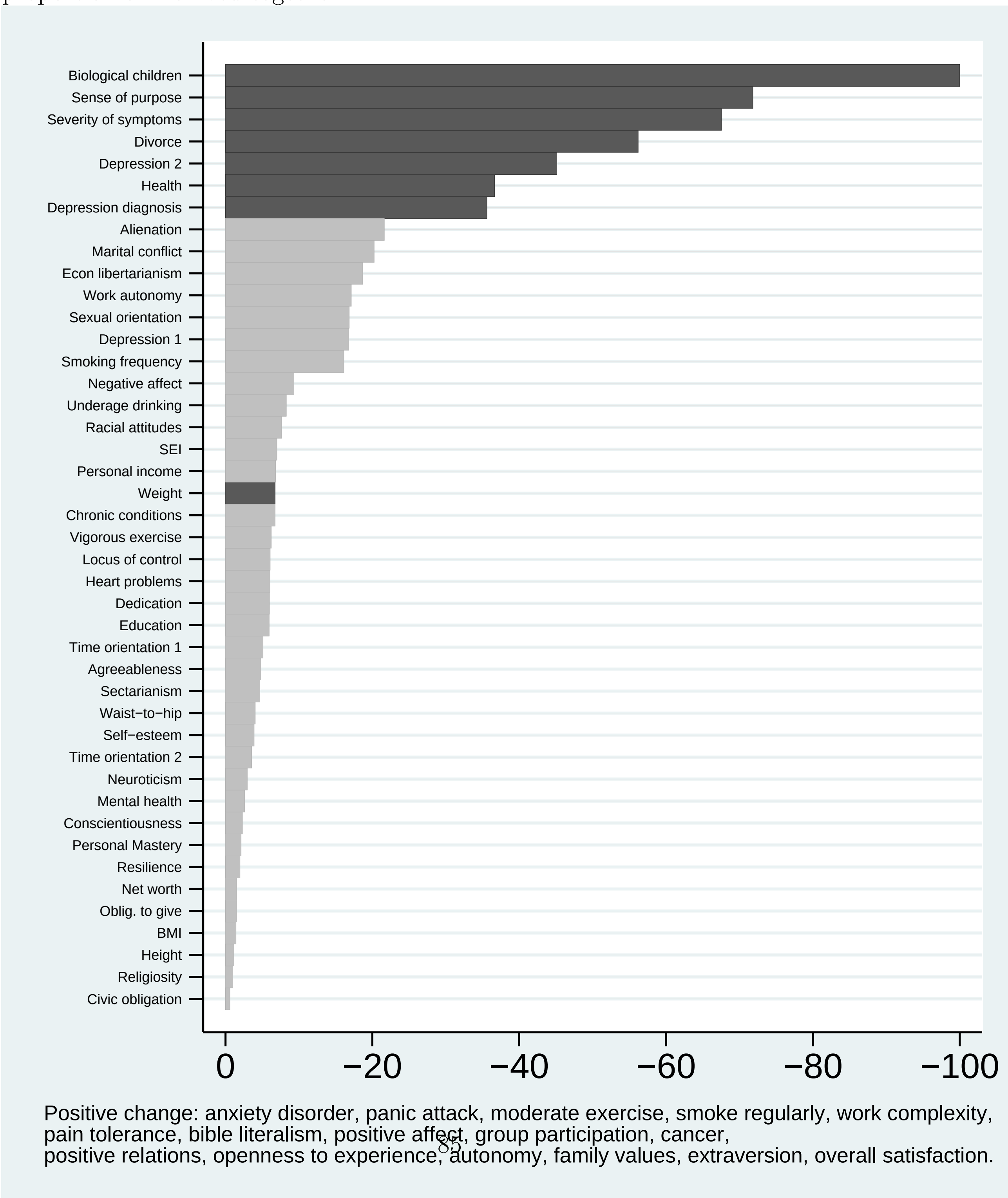


Figure 19: Percent change in the coefficient for zygosity when controlling for virtual contact

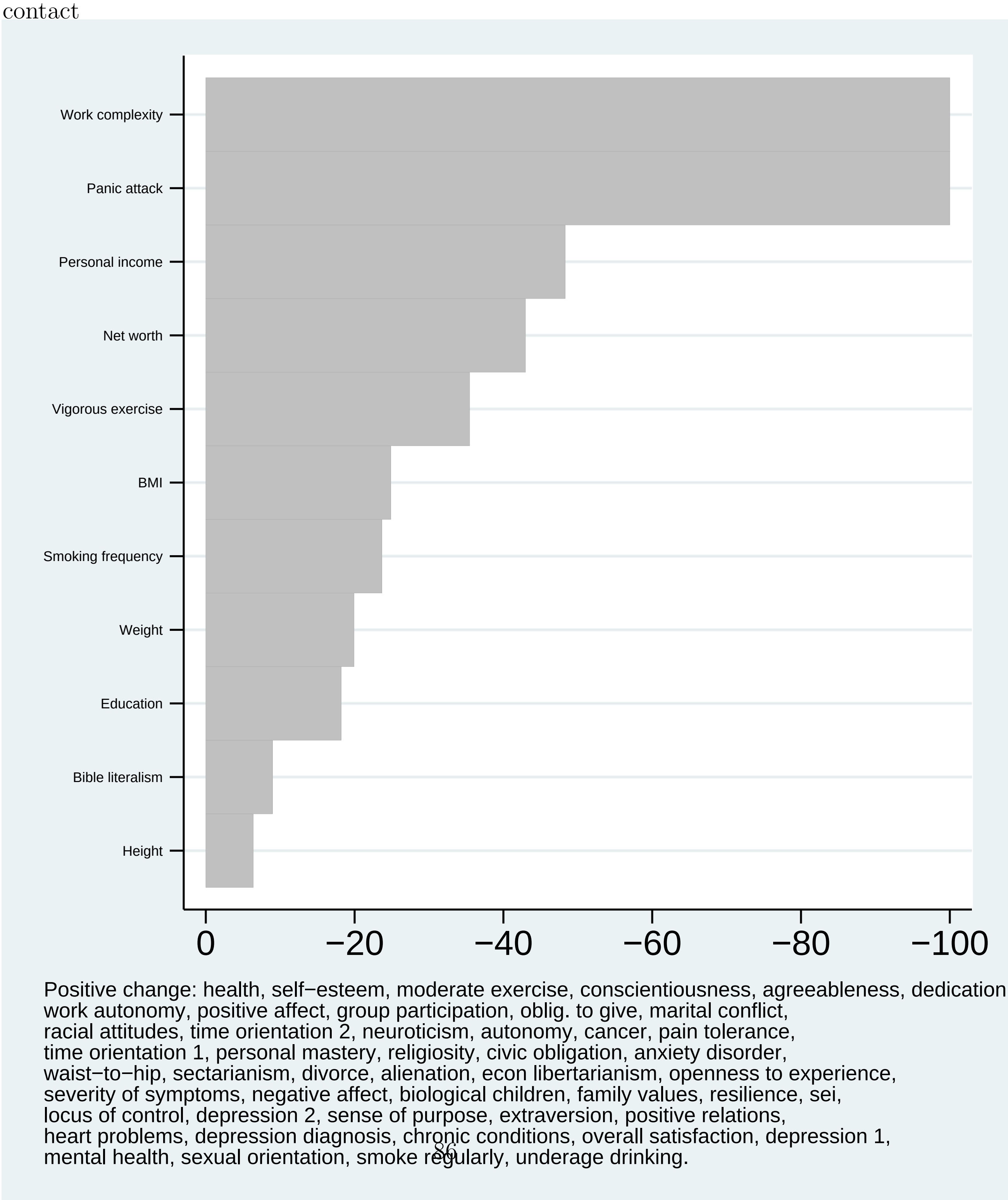


Figure 20: Percent change in the coefficient for zygosity when controlling for psychological intimacy

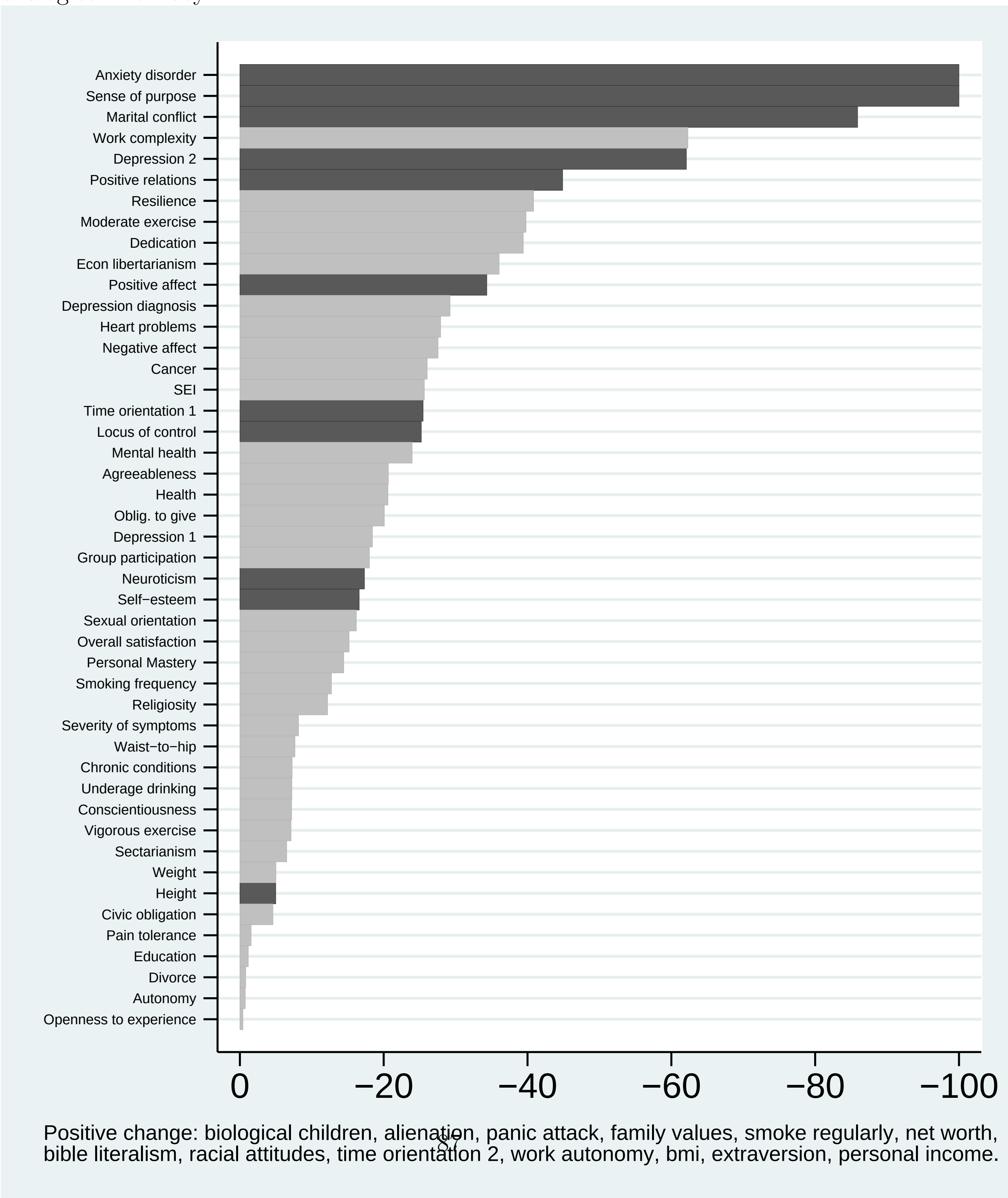


Figure 21: Percent change in the coefficient for zygosity when controlling for negativity in twin relationship

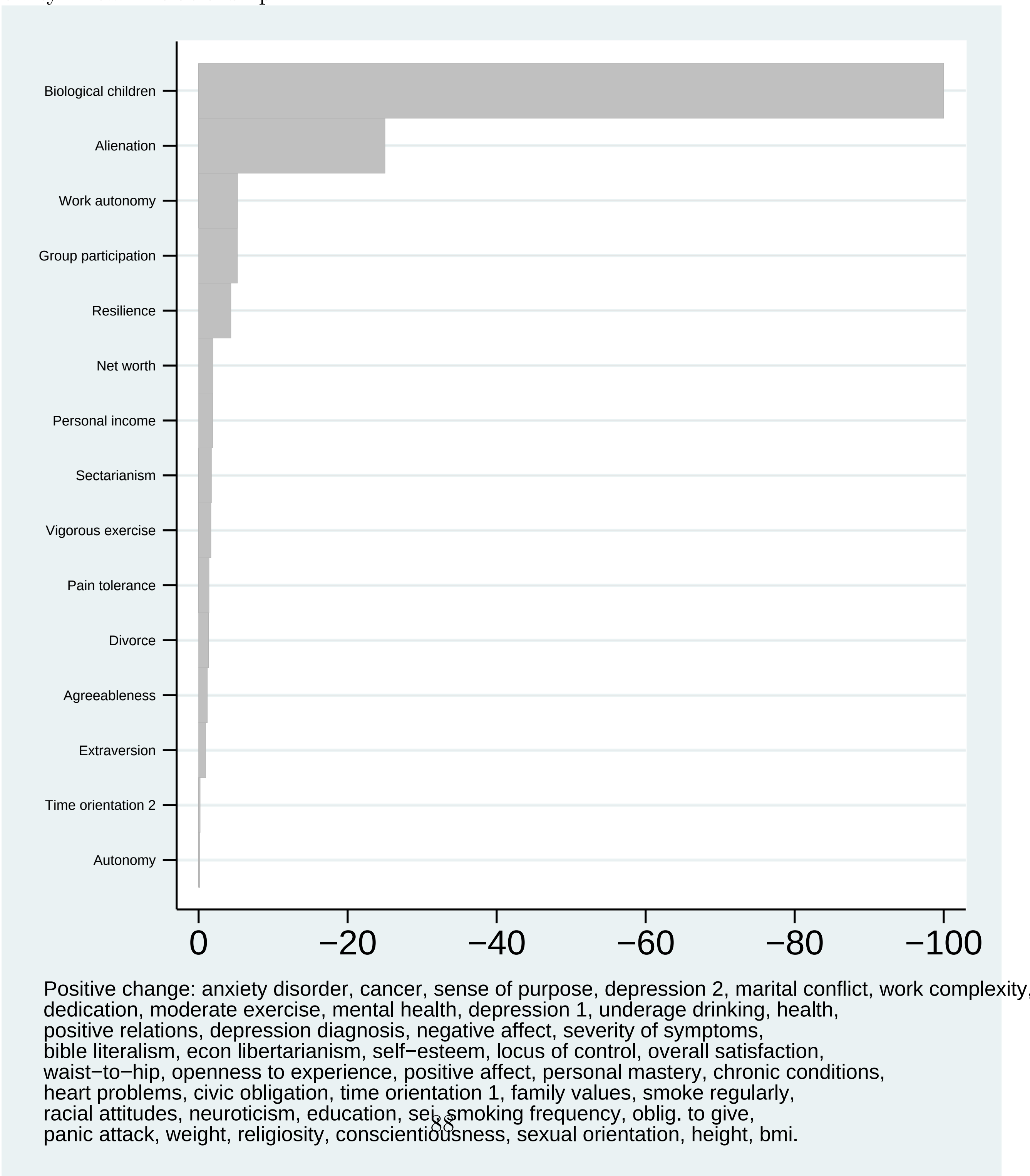


Figure 22: Percent change in the coefficient for zygosity when controlling for advice exchanged between twins

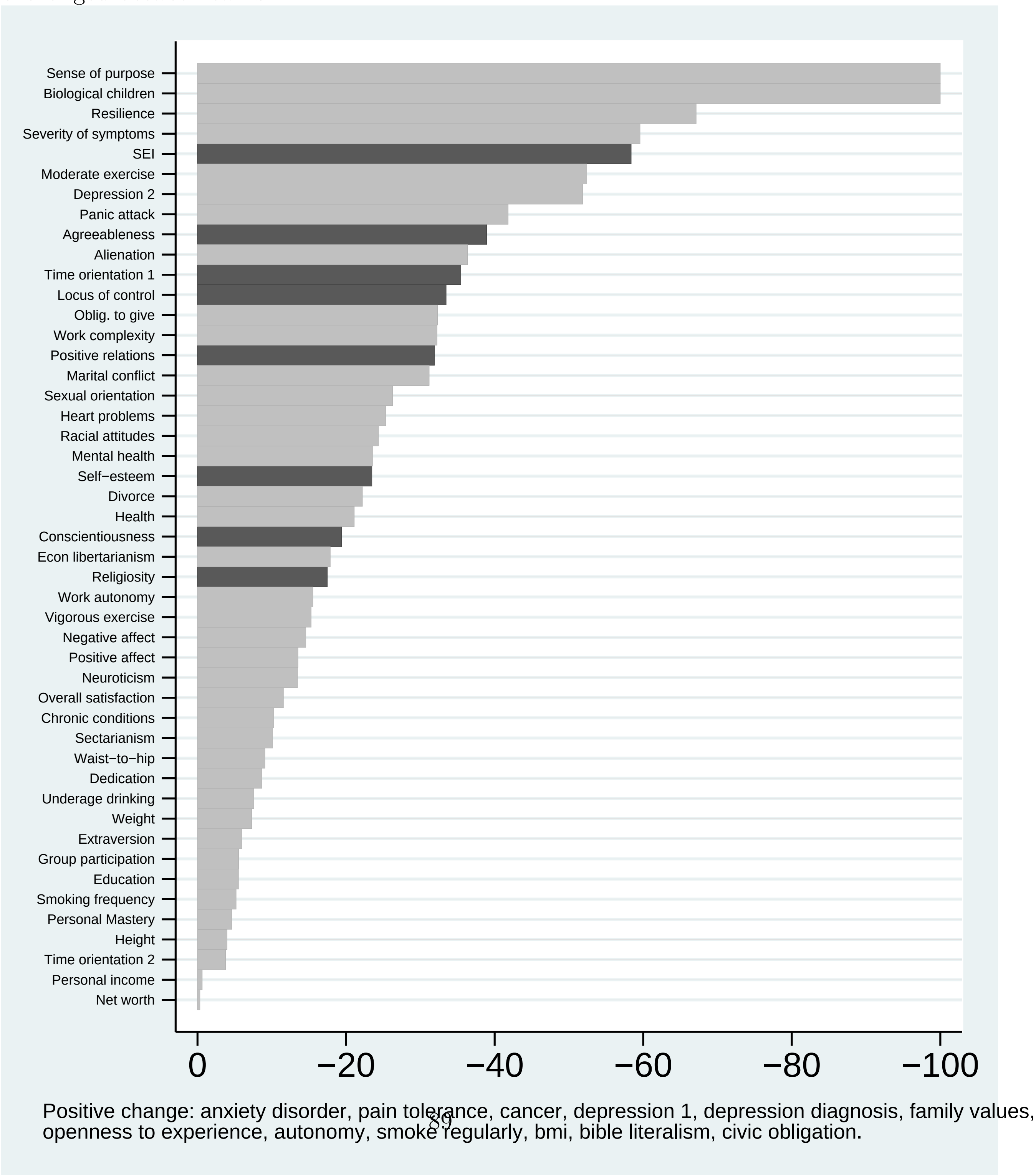


Figure 22 shows the percent change in the coefficient for zygosity when controlling for the extent of advice given and received between twins. As with previous measures of environmental similarity, there are few strong patterns here. One finding of note is that the apparent effect of genes on socioeconomic status appear to be confounded with the extent of advice exchanged between twins, as well as with childhood environmental similarity, and the amount of current contact. It may be the case that the equal environments assumption is most tenuous with regard to SEI. However, this conclusion is weakened by the fact that confounding is not found to the same extent with other measures of social class, such as income and net worth.

5.4 Defries-Fulker regressions

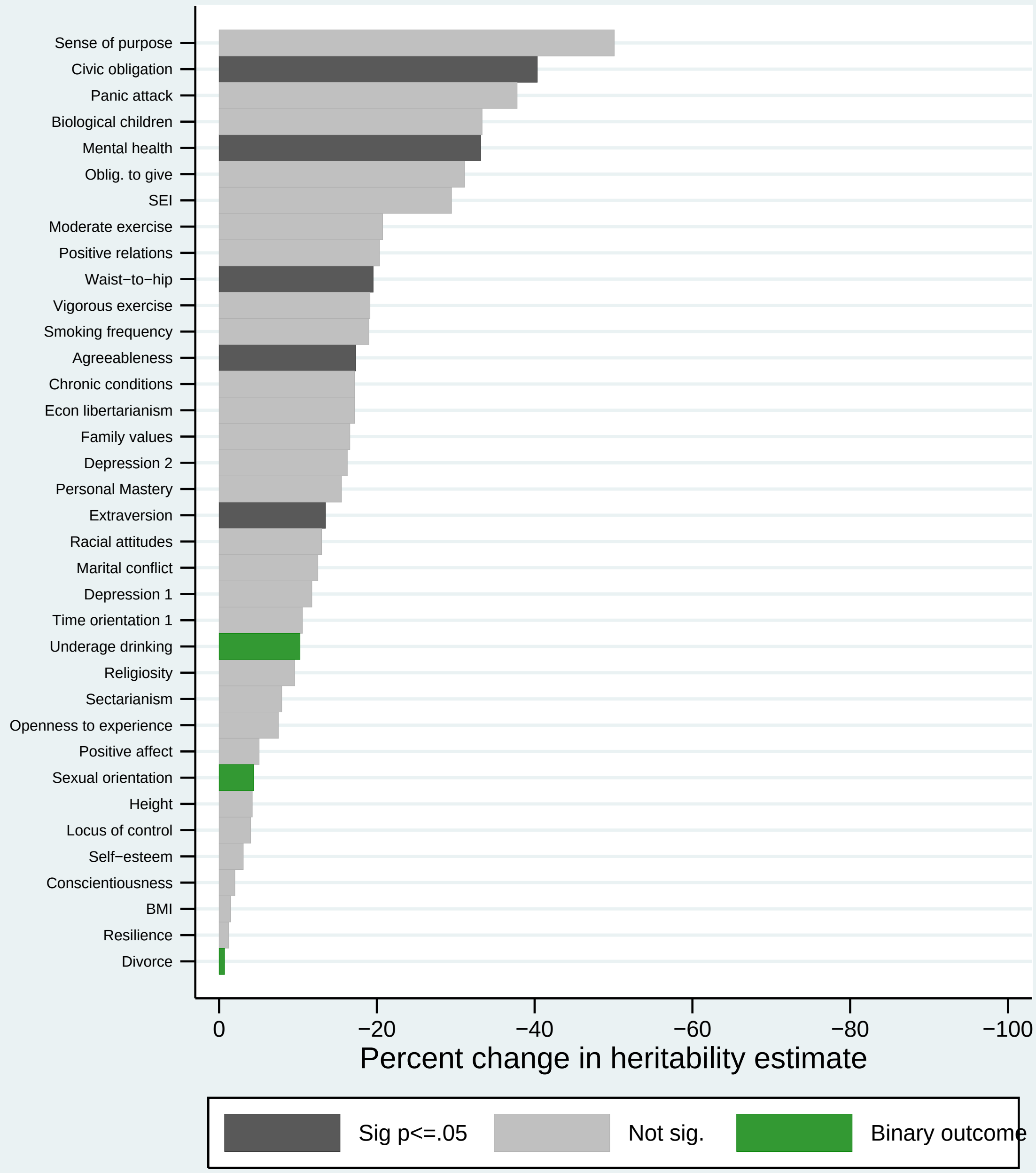
Figures 23-29 present results from DeFries-Fulker regressions with and without controls for each measure of environmental similarity. Again, the figures present only the most important results from these regressions: the percent change in the estimates of heritability when controlling for a particular measure of environmental similarity. In order to cut back on the size of these already dense figures, I omit outcomes for which environmental controls did not reduce the apparent effect of genes.

I am not aware of a method for evaluating the statistical significance of differences in coefficients across DeFries-Fulker-type regressions when the outcome is dichotomous. Thus, bars representing the confounding effects of environmental similarity for dichotomous outcomes were colored green as a reminder that no significance test was conducted. The bars representing results from regressions of non-dichotomous outcomes were shaded black if the confounding effect was significant.

Provided that the results of the previous set of regressions are not the product of chance fluctuations in the data, the results of those regressions should mirror the results in the DeFries-Fulker regressions. Surprisingly, however, there is little consistency between the results of the absolute value regressions and the results of the DeFries-Fulker regressions. While genetic effects on SEI appeared to be substantially weakened in the absolute value regressions, the genetic effects on SEI emerged unscathed in the DeFries-Fulker regressions.

In general, the equal environments assumption appears to be largely validated, if judged on the basis of DeFries-Fulker regressions in which one aspect of environmental similarity is controlled. The discrepancies between the results of the two methods raise the prospect that some of the significant results have arisen from chance fluctuations in the data that are bound to occur when so many regression coefficients are examined. The lack of a clear pattern of significance across categories of dependent variables reinforces this conclusion.

Figure 23: Percent reduction in heritability when controlling for childhood similarity



Positive change: group participation, work complexity, dedication, pain tolerance, time orientation 2, work autonomy, personal income, severity of symptoms, bible literalism, anxiety disorder, cancer, net worth, heart problems, alienation, neuroticism, autonomy, smoke regularly, overall satisfaction, negative affect, depression diagnosis, weight, health, education.

Figure 24: Percent reduction in heritability when controlling for inequality in twin relationship growing up

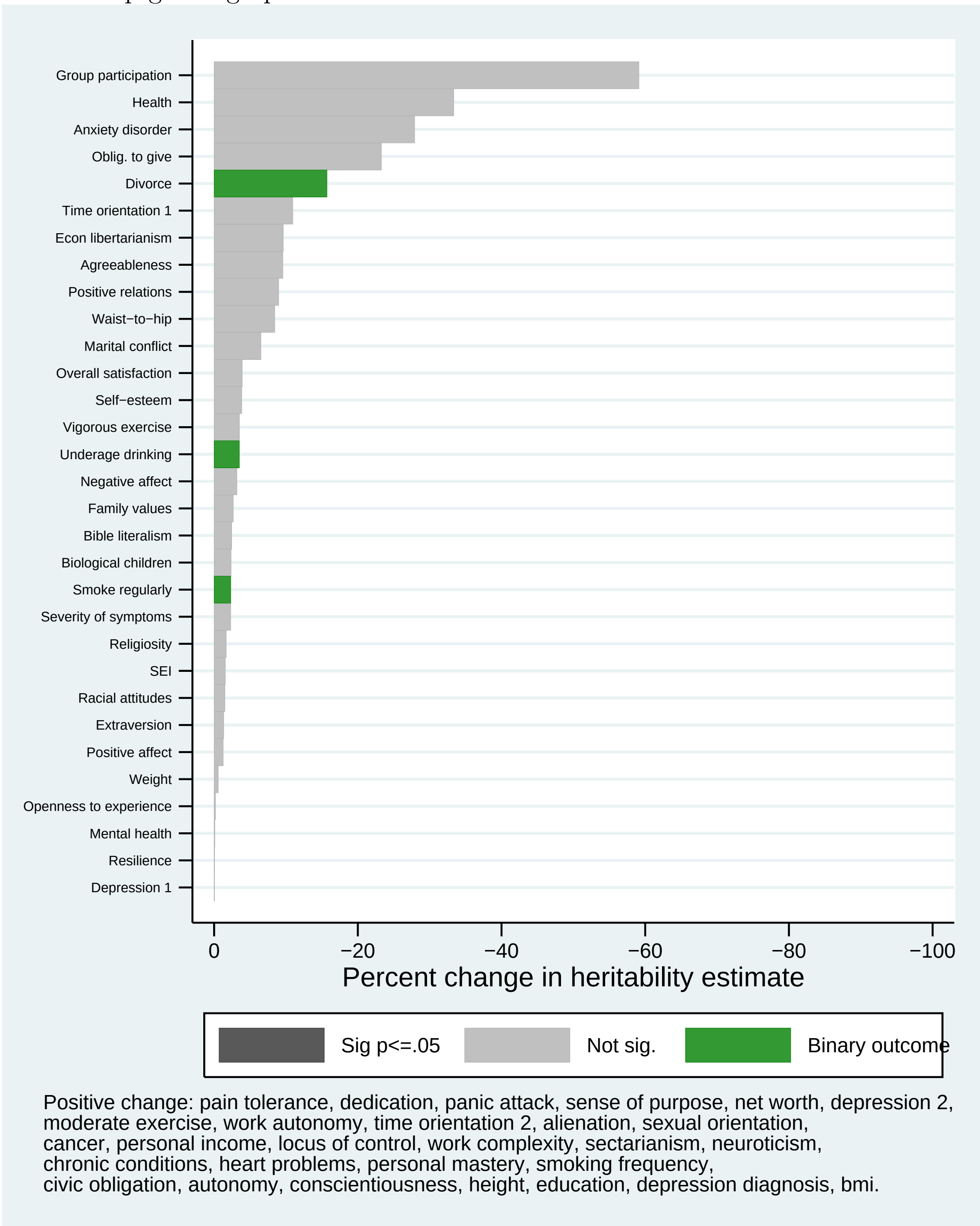


Figure 25: Percent reduction in heritability when controlling for current contact

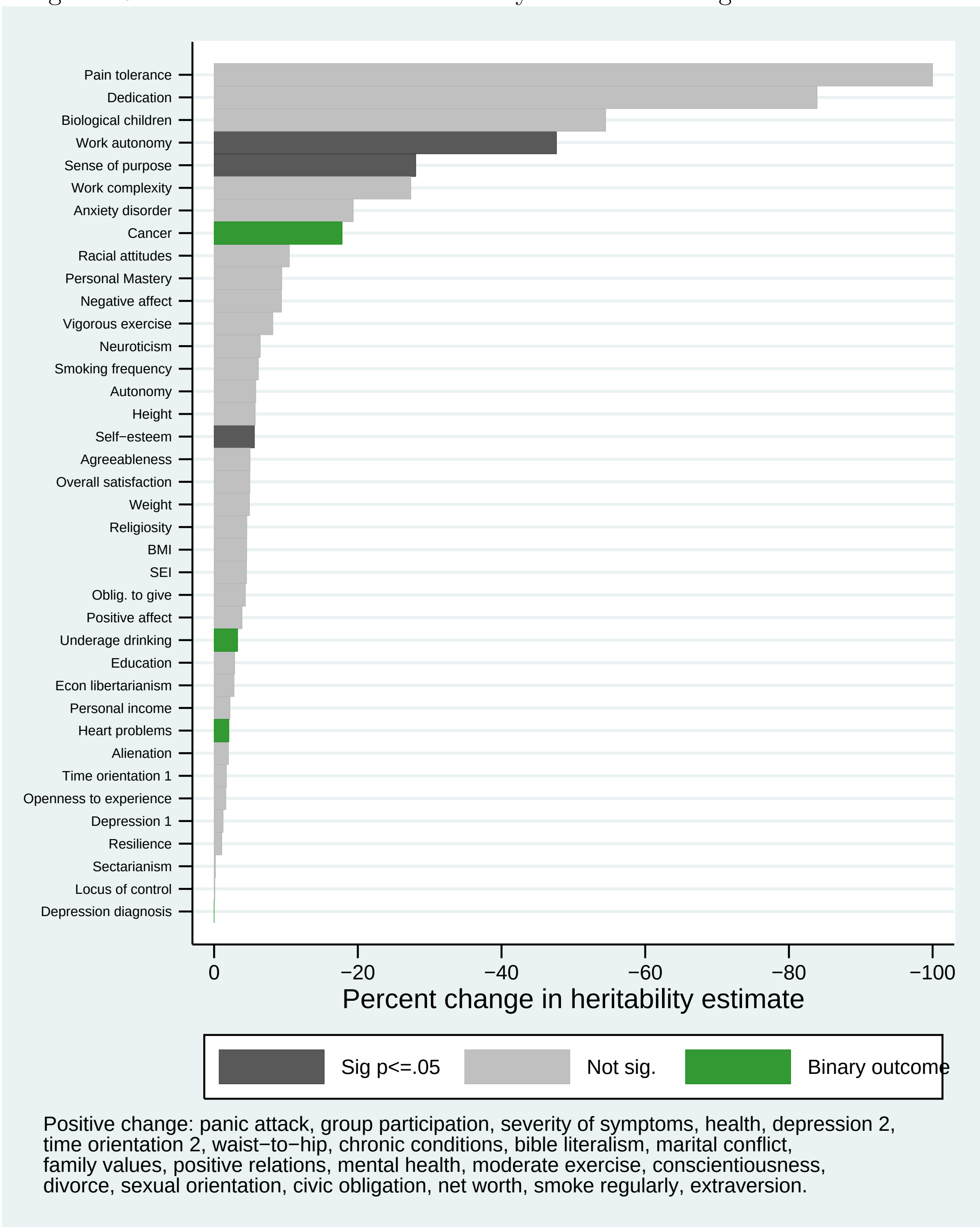
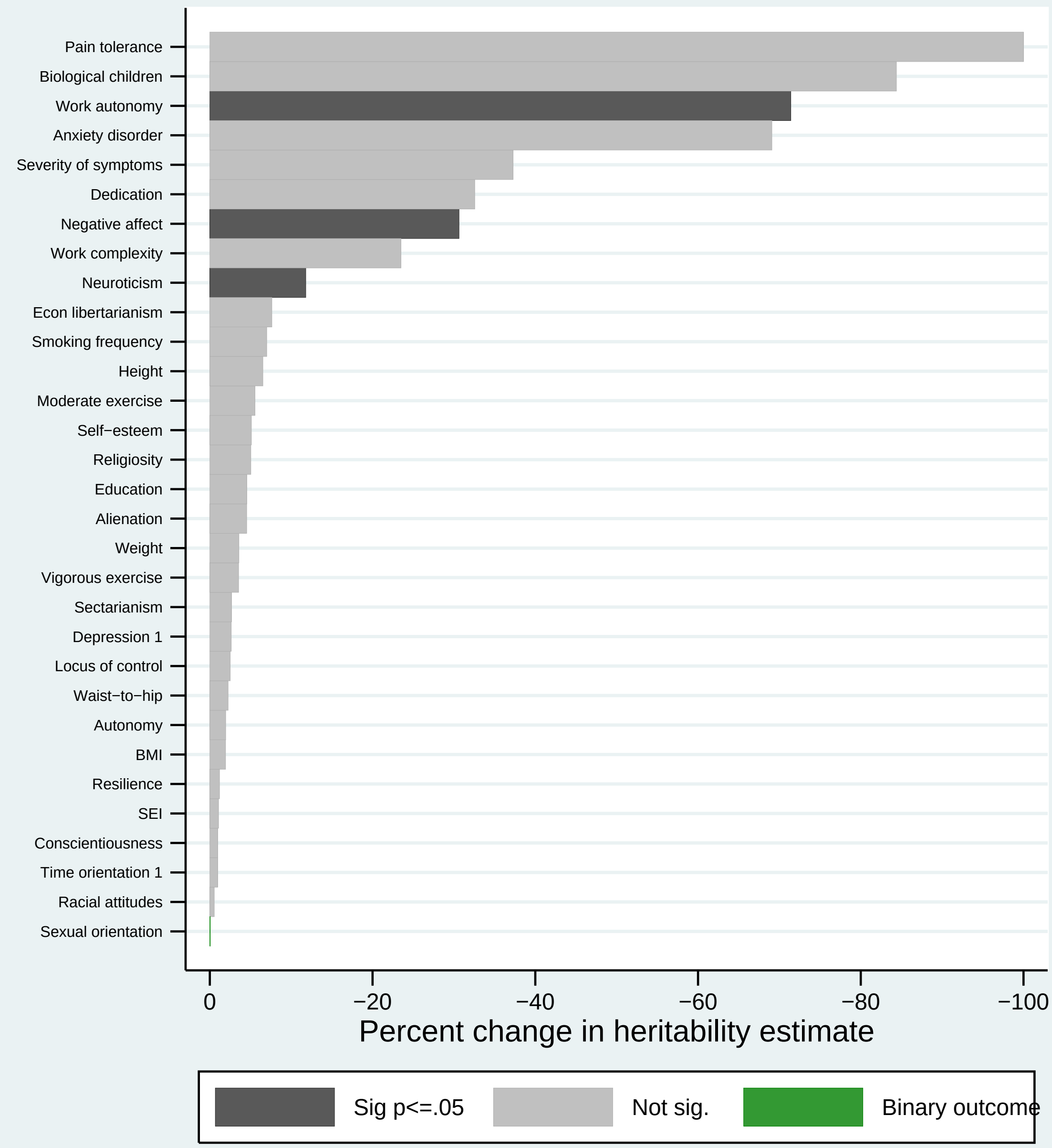
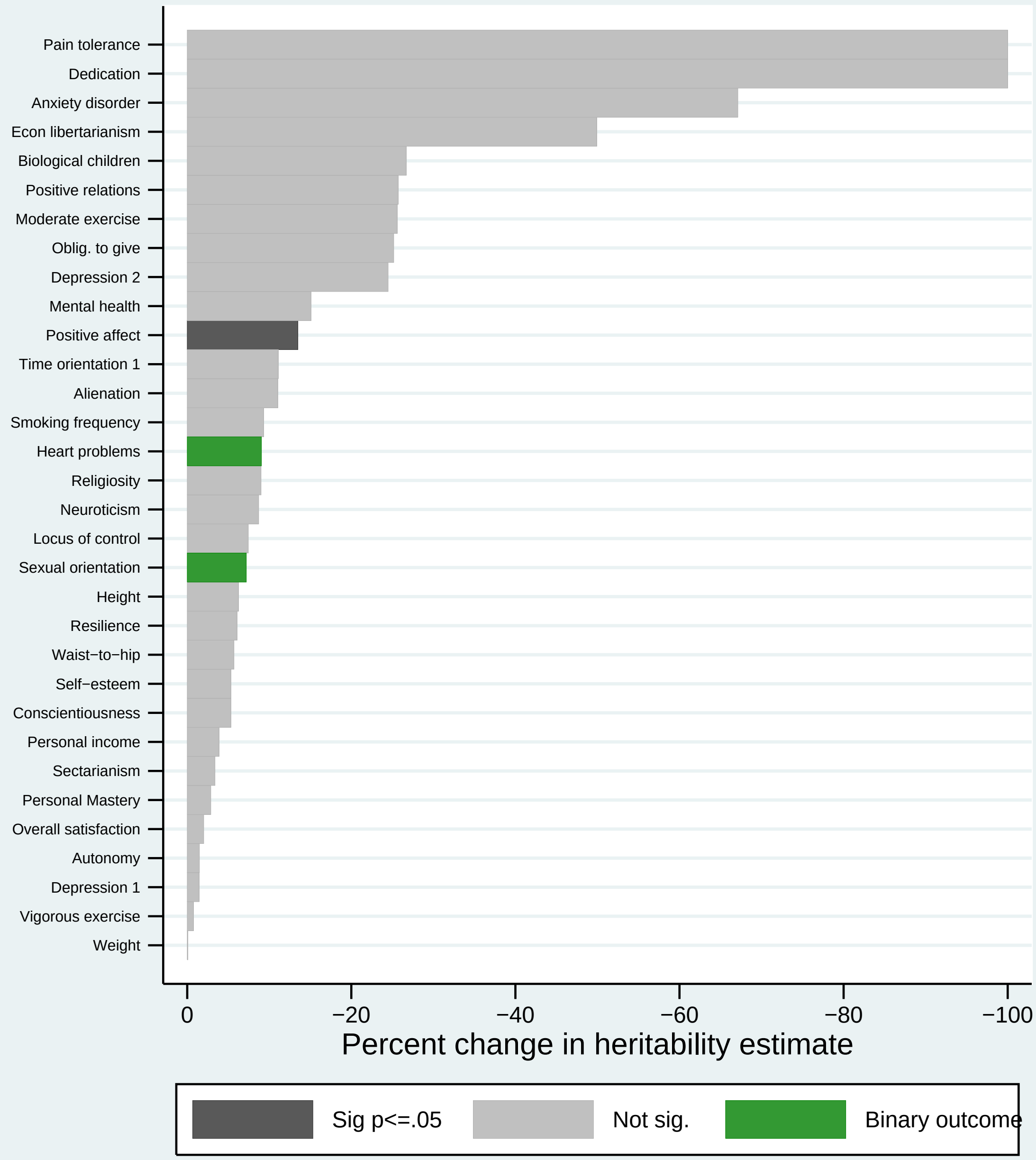


Figure 26: Percent reduction in heritability when controlling for proportion of life spent together



Positive change: panic attack, health, divorce, personal income, heart problems, cancer, chronic conditions, group participation, oblig. to give, marital conflict, depression 2, bible literalism, time orientation 2, underage drinking, mental health, net worth, positive relations, civic obligation, sense of purpose, depression diagnosis, extraversion, openness to experience, personal mastery, family values, smoke regularly, agreeableness, overall satisfaction, positive affect.

Figure 27: Percent change in heritability when controlling for psychological intimacy



Positive change: group participation, work complexity, panic attack, family values, severity of symptoms, divorce, time orientation 2, cancer, net worth, bible literalism, civic obligation, health, negative affect, sei, chronic conditions, bmi, work autonomy, depression diagnosis, smoke regularly, racial attitudes, agreeableness, extraversion, marital conflict, education, openness to experience, underage drinking, sense of purpose.

Figure 28: Percent change in heritability when controlling for negativity in twin relationship

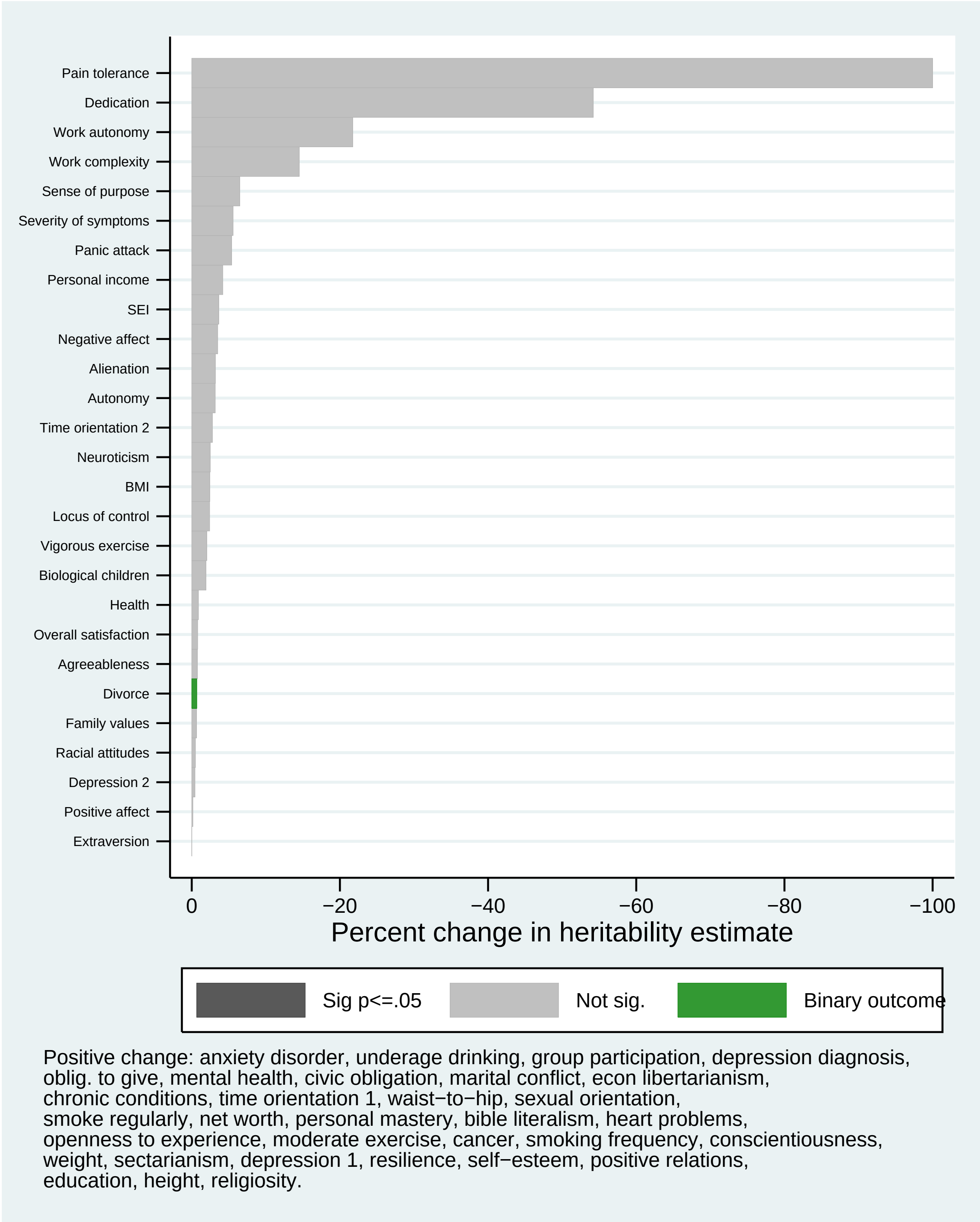
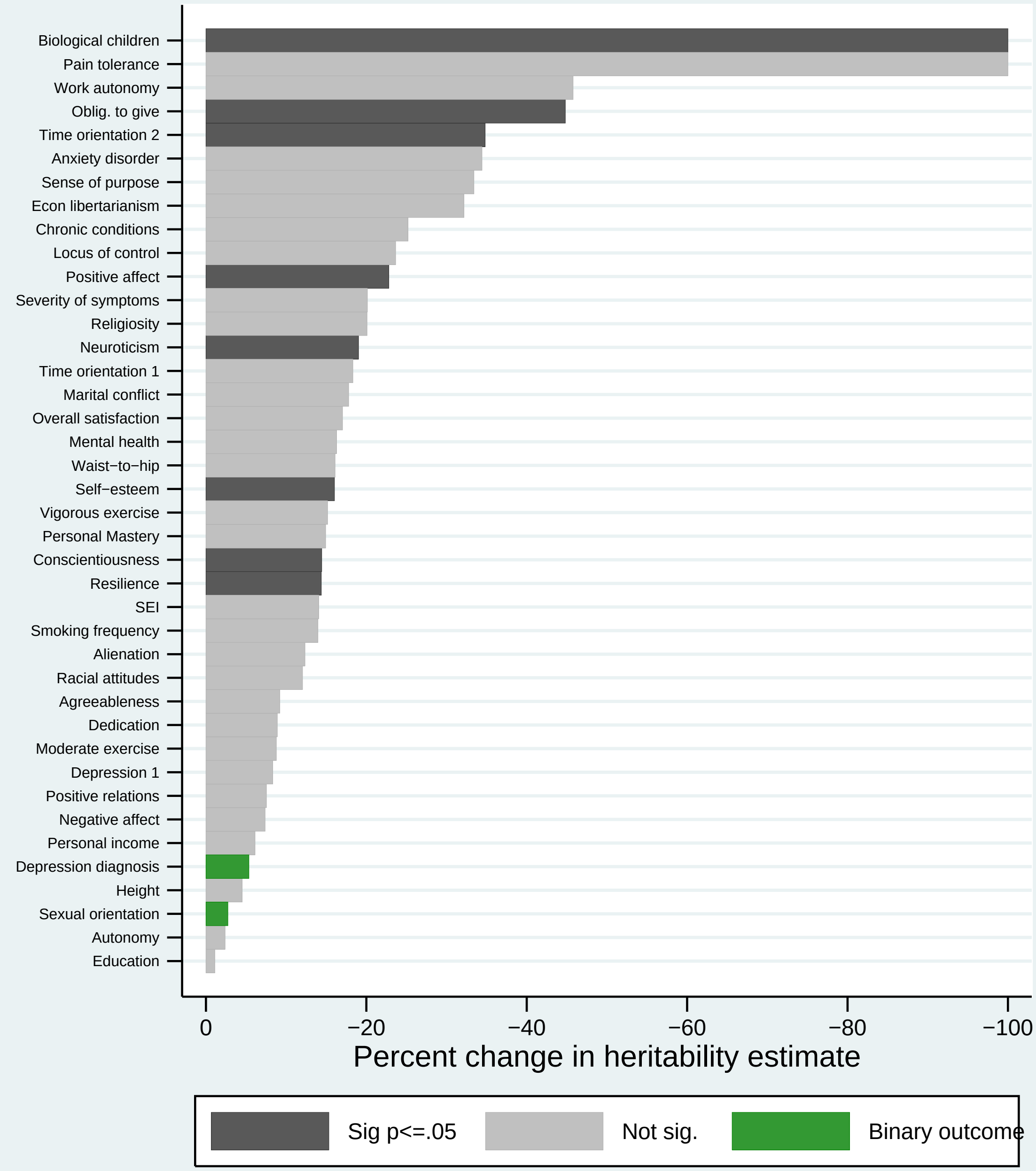


Figure 29: Percent change in heritability when controlling for advice exchanged between co-twins

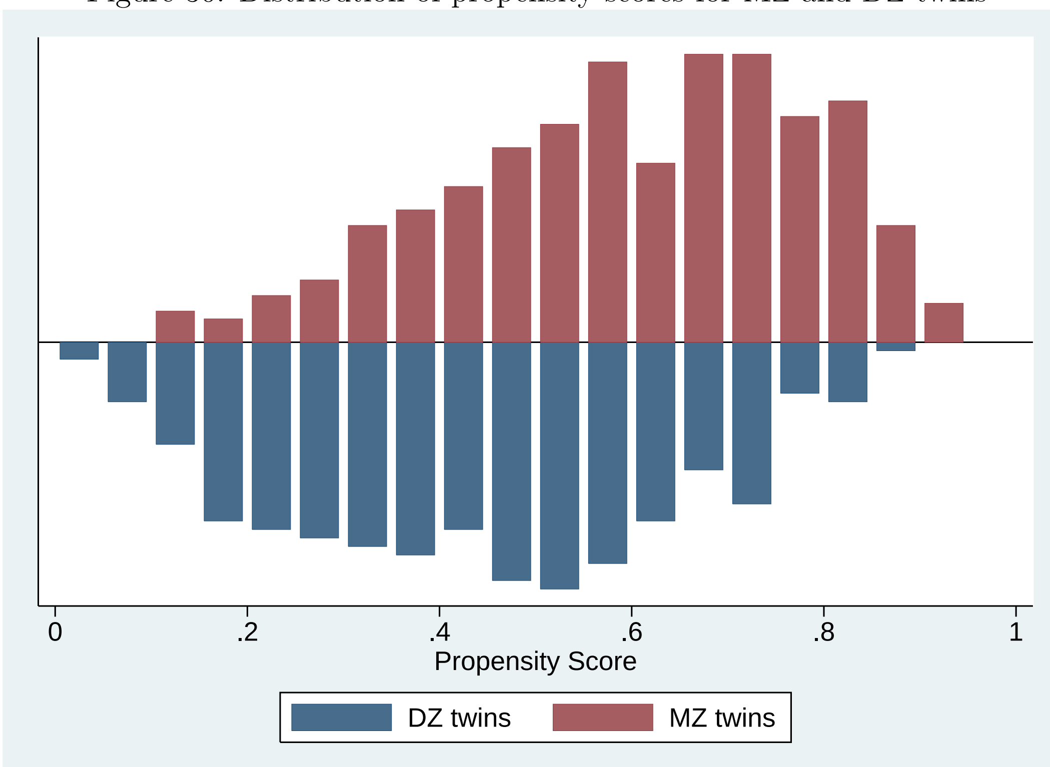


Positive change: health, panic attack, group participation, family values, depression 2, work complexity, net worth, bible literalism, bmi, underage drinking, smoke regularly, openness to experience, extraversion, cancer, civic obligation, heart problems, divorce, weight, sectarianism.

5.5 Propensity score matching

Figures 30-44 present the results of the final set of analyses, in which I look at the effect of controlling for all aspects of similarity in the environments reported by the twins in MIDUS. Figure 30 presents propensity scores separately by zygosity. As explained earlier, propensity scores are predicted probabilities generated from a logistic regression of monozygosity on all measures of environmental similarity. The bars for MZ twins point upwards, while the bars for DZ twins point downwards.

Figure 30: Distribution of propensity scores for MZ and DZ twins



There is considerable overlap between the propensity score distributions for each twin type. This is good news; there appear to be a substantial number of DZ twins who experienced environments as similar as those experienced by MZ twins, at least as far as we can tell with these measures. To be sure, there are MZ twins with propensity scores unmatched by DZ twins, and vice-versa. It is unclear whether a larger sample would pick up MZ twins (reared together) whose environments are as dissimilar as the most disparate DZ twins, or analogously, whether a larger sample would pick up DZ twins whose environments are as similar as that of MZ twins who were most environmentally concordant. It is possible that no DZ twins experience environments as similar as the most environmentally similar MZs, and that no MZs experience environments as dissimilar as the most dissimilar DZs.

At any rate, the propensity score distributions in Figure 30 clarify the comparison at the heart of the debate about EEA. In effect, one tries to find DZ twin pairs who appear to have been as close – both physically and psychologically – to each other as many MZ twin pairs. A methodologically rigorous twin study must compare only those MZ and DZ twins who share the same place along the x-axis in Figure 30.

Figure 31: Balance on measures of environmental similarity

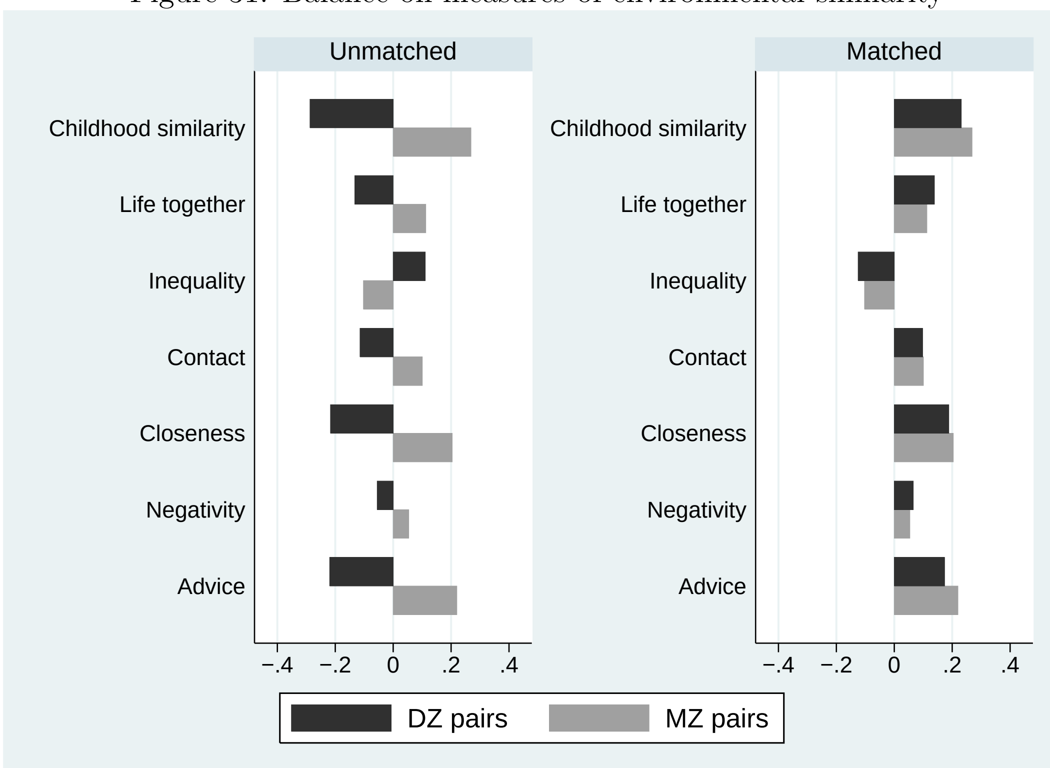


Figure 31 shows the differences between MZ and DZ twins in terms of environmental similarity before and after matching. All seven measures of environmental similarity are measured in terms of standard deviations.

Prior to matching, one can see that the zygosity differences are largest with respect to the similarity of treatment in childhood, and with respect to the extent of advice exchanged. By contrast, MZ and DZ twins differ very little in terms of the amount of negativity perceived in the relationship.

Zygosity differences on all of these measures become insignificant in the matched data. The differences shown in the bar chart on the right are very small. In addition, t-tests for each measure revealed no significant differences on any measure in the matched data.

Figures 32-44 show comparisons of heritability estimates with and without matching. Each figure displays heritability estimates for one of the thirteen categories of outcomes.

Figure 32: Heritability estimates for physical measurements

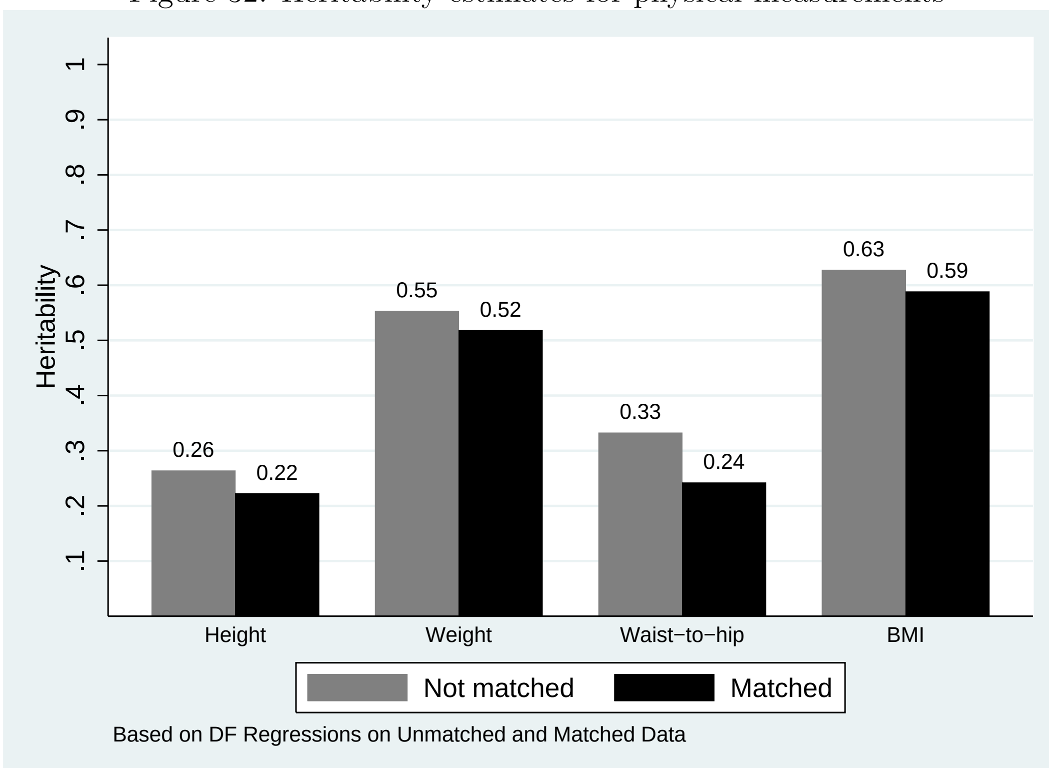
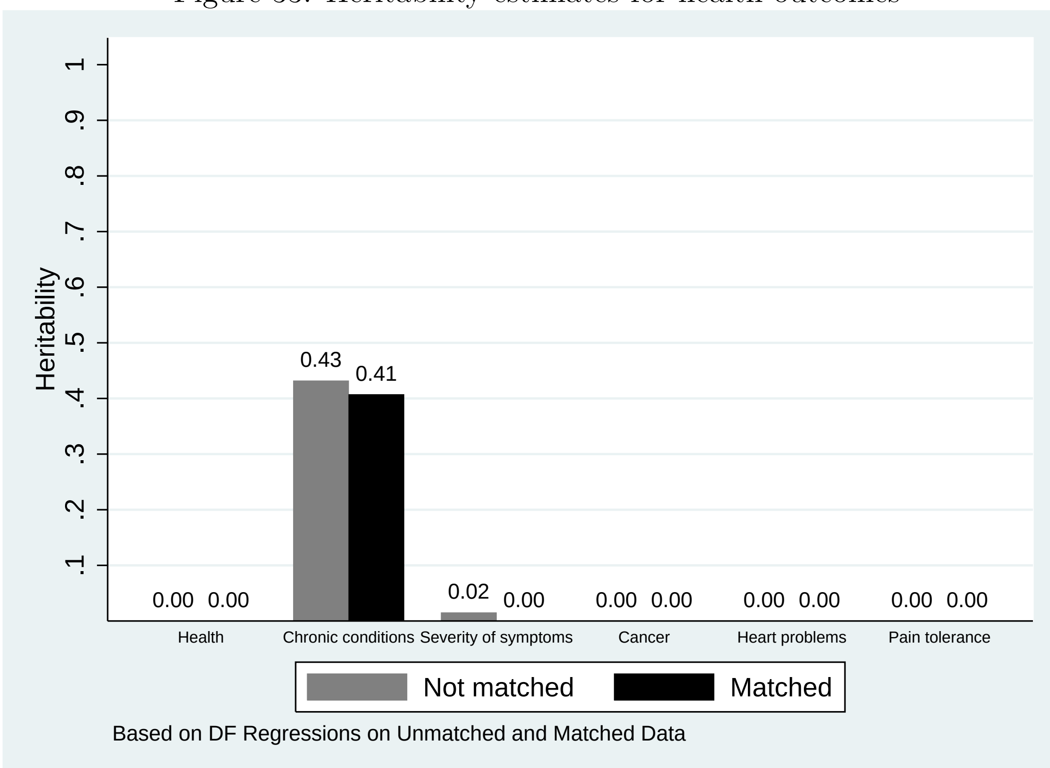


Figure 32 shows heritability estimates for physical characteristics. Unsurprisingly, the effect of matching on environmental similarity on the heritability of physical characteristics is small.

Figure 33 shows heritability estimates for health outcomes. By chance, MIDUS did not reveal any genetic effect for several of these variables, so I am stymied in my ability to find any confounding influence of the environment. The zero heritability for health may be due to the fact that the population in MIDUS is fairly young and thus generally healthy. Since there isn't likely to be great variation in health in the

Figure 33: Heritability estimates for health outcomes



MIDUS sample, there would be little variation to explain by genes. The non-zero heritabilities for heart problems and for number of chronic conditions are robust to controls for the environment. This is expected; most skeptics of twin studies have focused their criticism on behavior geneticists' analysis of social and psychological traits.

Figure 34: Heritability estimates for health-related behaviors

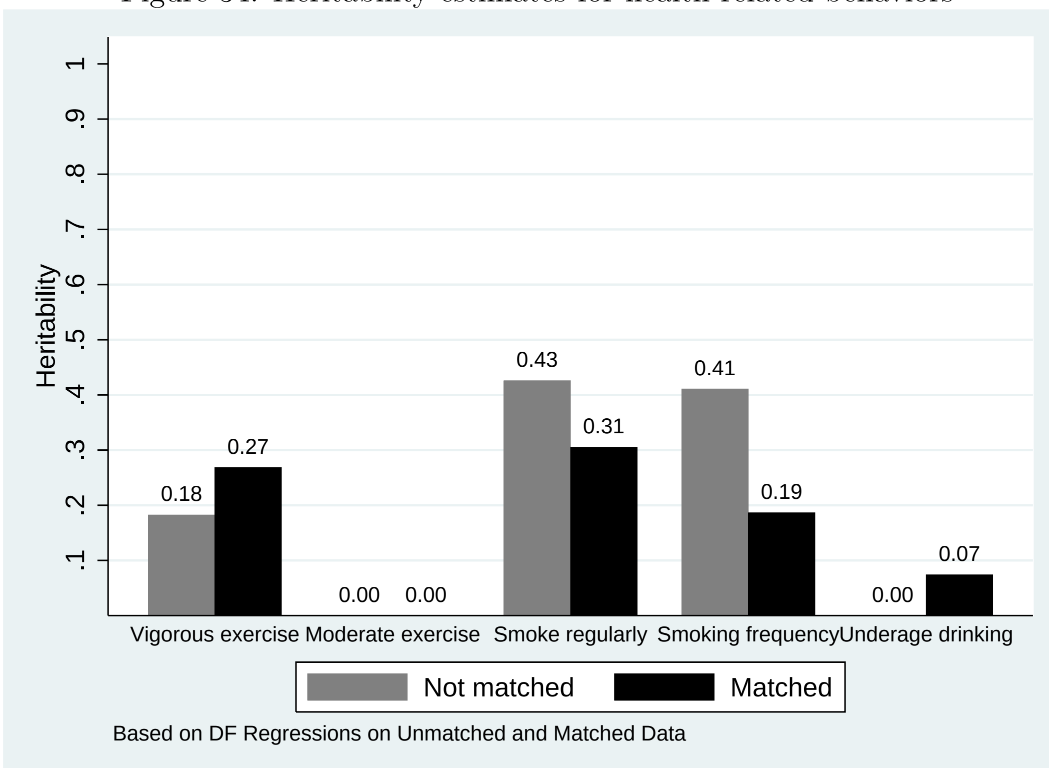
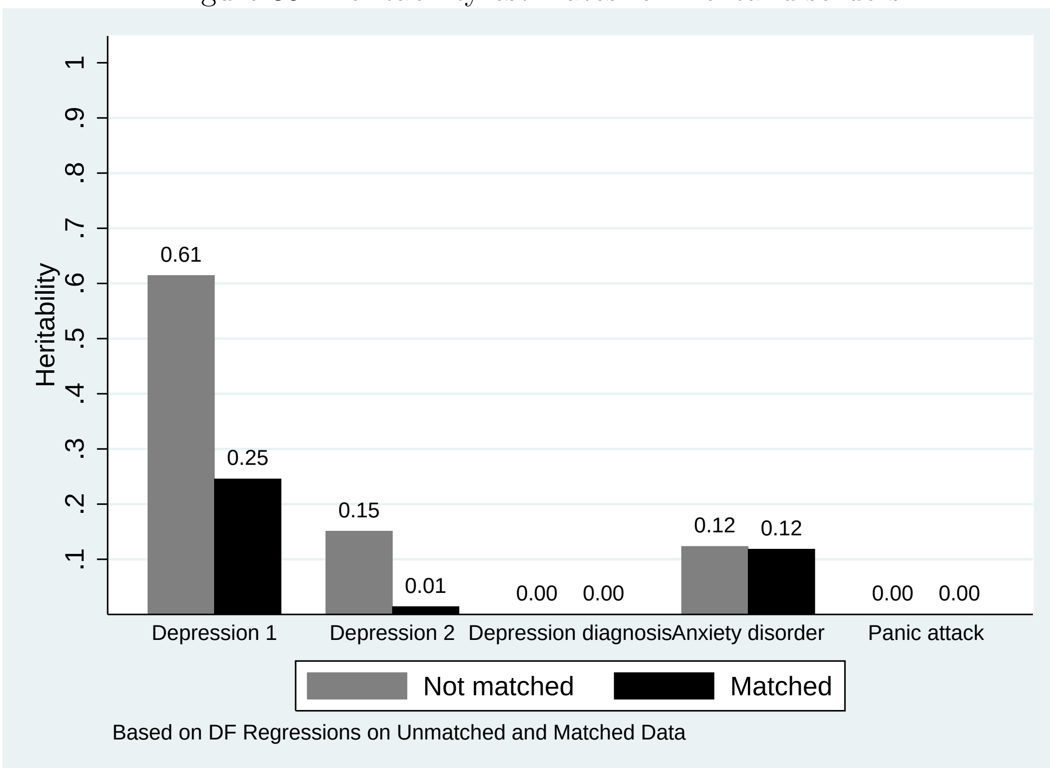


Figure 34 shows heritability estimates for behaviors that have a strong impact on health. Note that the outcome "smoke regularly" is a dichotomous measure of whether the person ever smoked regularly in their life. The heritability estimate associated with this variable is indirect; heritability here is an estimate of the genetic influence on the propensity to begin smoking regularly, a latent variable that is measured only indirectly by reports of whether a person ever smoked regularly.

While heritability estimates on propensity to smoke are robust to environmental controls, heritability estimates for smoking frequency drop by half on the matched data. It seems imperative that twin studies of smoking include controls for environmental similarity.

Figure 35: Heritability estimates for mental disorders



A similar story emerges with respect to measures of depression in Figure 35. As with the heritability estimate for "smoking regularly," the heritability estimate for "depression diagnosis" reflects the genetic influence on the latent propensity to be diagnosed as depressed. Again, this (theoretical) latent trait appears robust in the matched data, while heritability estimates for similar, non-latent measures decline dramatically in the matched data. It is unclear whether these differences reflect substantive differences in the interplay of nature and nurture, or whether they arise from the different methods used to obtain heritability estimates for dichotomous and non-dichotomous data.

Figure 36: Heritability estimates for personality characteristics

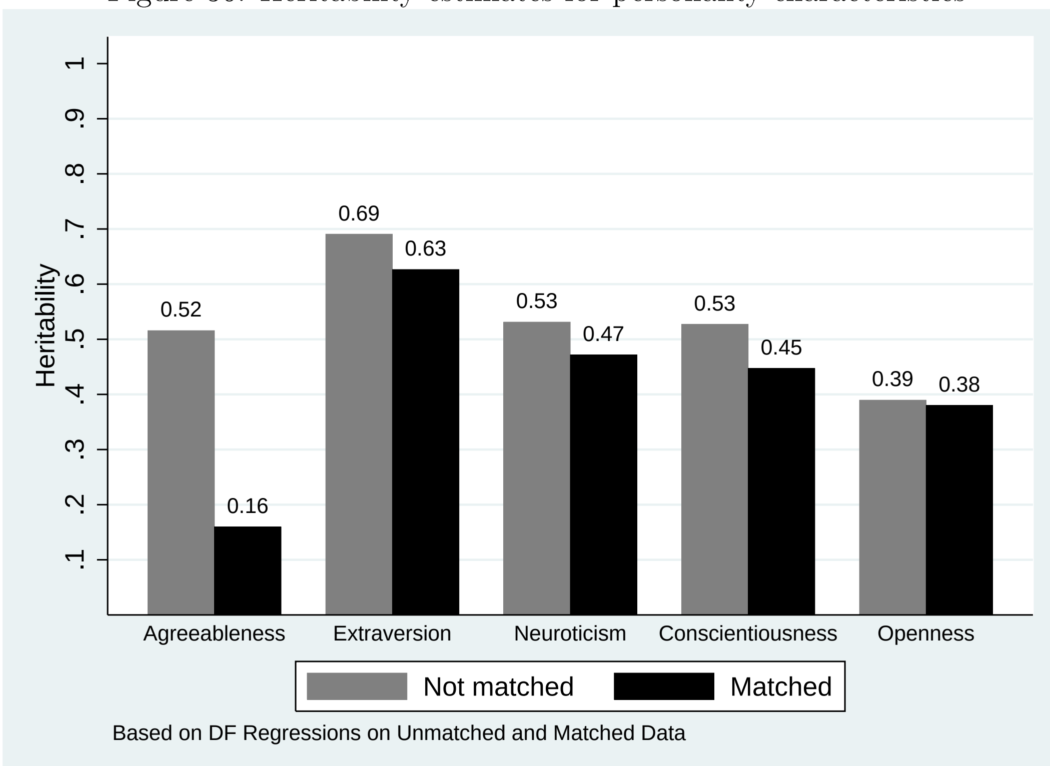
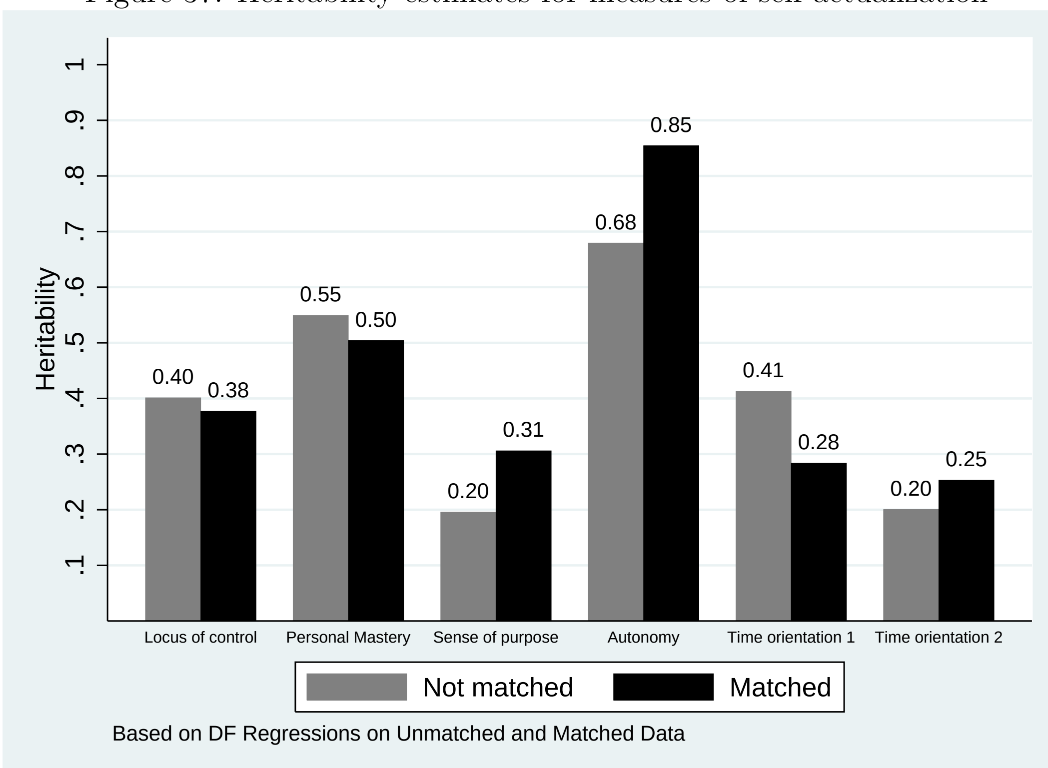


Figure 36 reveals that heritability estimates for four out of the five major personality characteristics identified by psychologists appear robust to controls for environmental similarity. By contrast, estimated heritability for agreeableness plummets by more than 50%.

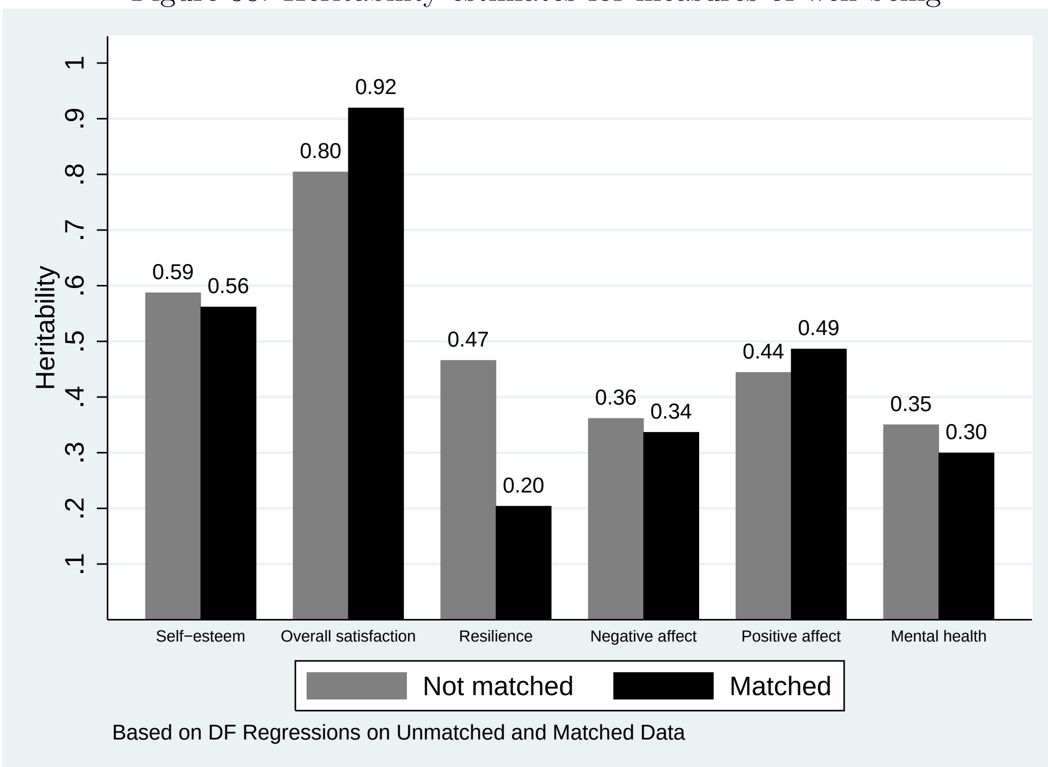
There is little evidence in Figure 37 that estimates of heritability are compromised by violations of the equal environments assumption. The same conclusion applies to

Figure 37: Heritability estimates for measures of self-actualization



most heritability estimates for measures of well-being in Figure 38. The exception is the estimate of heritability for resilience, which drops by more than half. However, this result should not form the basis of strong conclusions; there is no pattern in the heritability estimates for measures of well-being.

Figure 38: Heritability estimates for measures of well-being

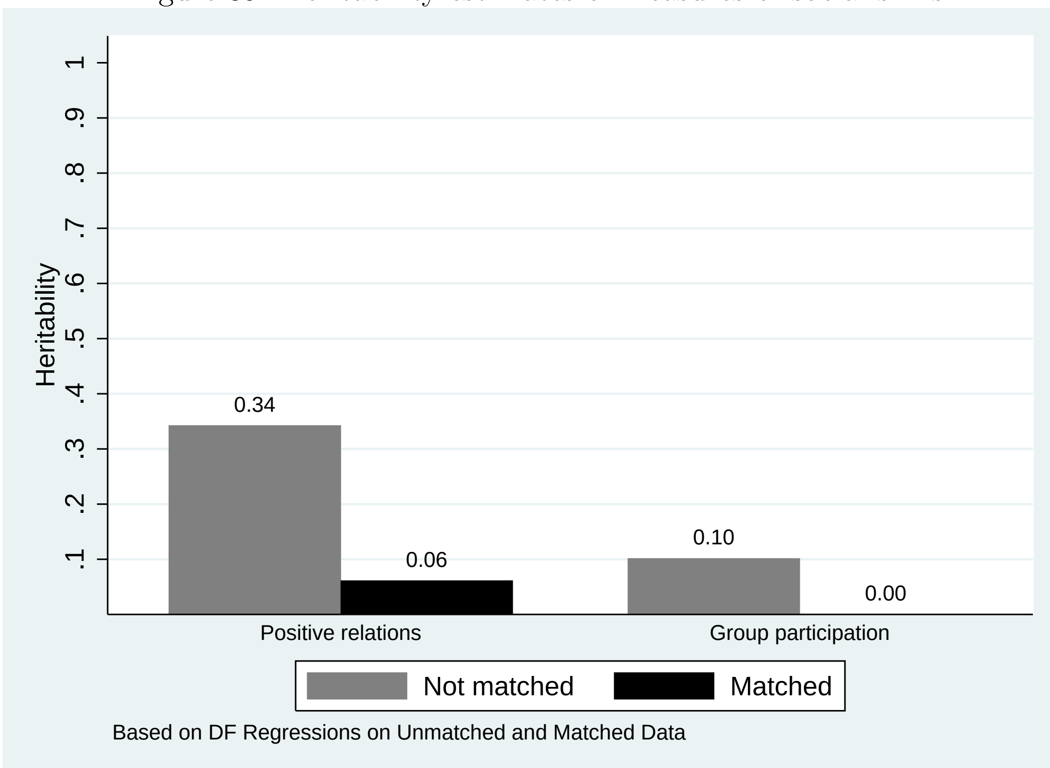


The results for social relations, displayed in Figure 39, are a different story. Here we see evidence for substantial confounding with both measures of social relations – the extent of positive relations with others (as reported by the twins themselves) and the extent of group involvement.

We find mixed support for the equal environments assumption in Figure 40. Consistent with regression analyses presented earlier, the heritability estimate for fertility vanishes in the matched data. This could raise important questions for research in this area by demographers such as Kohler. Previous work on the equal environments assumption has not looked at whether this assumption is valid in twin studies on fertility.

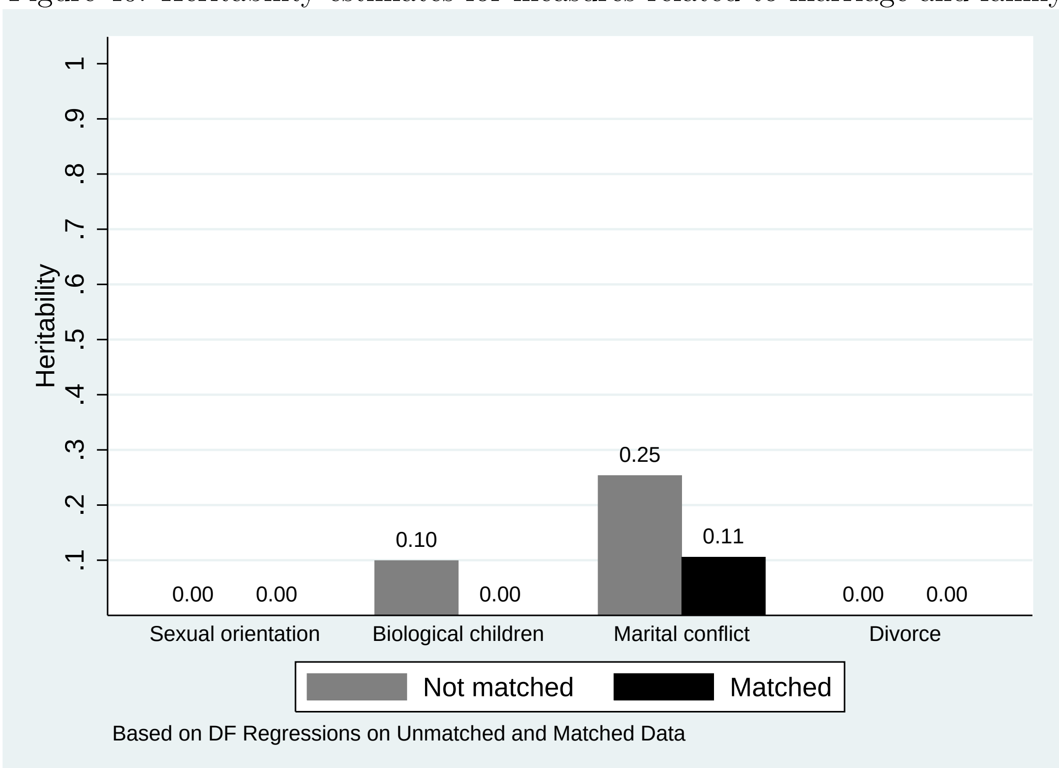
There appears to be an inconsistency between the non-robustness of heritabil-

Figure 39: Heritability estimates of measures of social skills



ity for marital conflict, and the robustness of the estimate of heritability for the divorce measure. Again, it is important to consider that the divorce measure is dichotomous, and hence the method for calculating heritability is rather different. The same method was used for the dichotomous measure of sexual orientation. It is notable that heritability estimates for all dichotomous measures have been invariant to matching. As mentioned earlier, this result could be a methodological artifact.

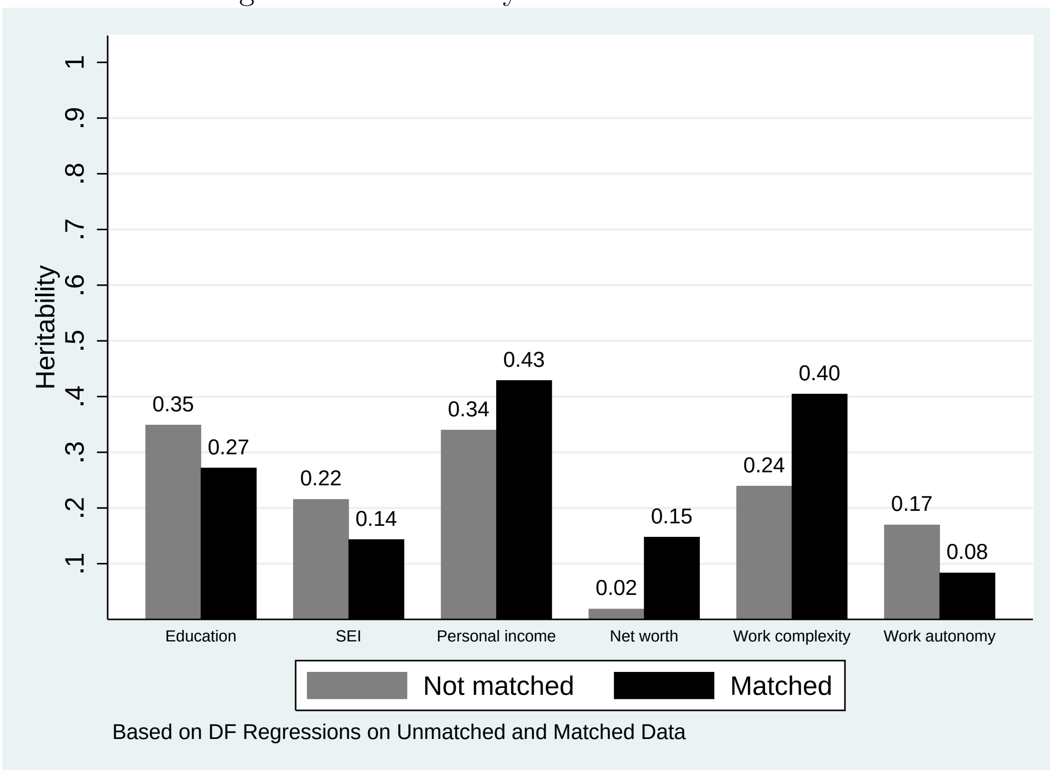
Figure 40: Heritability estimates for measures related to marriage and family



Perhaps Figure 41 displays the results that most interest sociologists. Given the inconsistency in the changes in the heritability estimates between unmatched and matched data, it is difficult to reach definitive conclusions. However, given the prodigious efforts to control for environmental similarity as thoroughly as possible, the equal environments assumption appears surprisingly robust.

Surprisingly, the apparent effect of genes on socio-political attitudes, such as family values, racial attitudes and alienation, appears unconfounded with environmental similarity. I cannot think of a substantive reason why heritability estimates would increase when controlling for environmental similarity. These results suggest that the

Figure 41: Heritability estimates for social class



chief problem with the twin study may not be the equal environments assumption, but rather, a lack of reliability. The sensitivity of the results to model specification suggests that single-number heritability estimates may be misleading. That said, heritability estimates are almost invariably greater than zero, using matched as well as unmatched data.

In Figure 43, we see that the genetic effect on perceptions of civic obligation and responsibility vanishes when MZ and DZ twins are matched with respect to environmental similarity. This makes sense because it is likely that perceived civic obligation is heavily influenced by social forces.

The results for religion are presented in Figure 44. The near-doubling of the estimate for heritability of beliefs about the bible seems to be a statistical fluke, and illustrates the volatility of heritability estimates. Heritability estimates for religiosity and sectarian religious beliefs fall rather modestly in the matched data.

Figure 42: Heritability estimates for socio-political attitudes

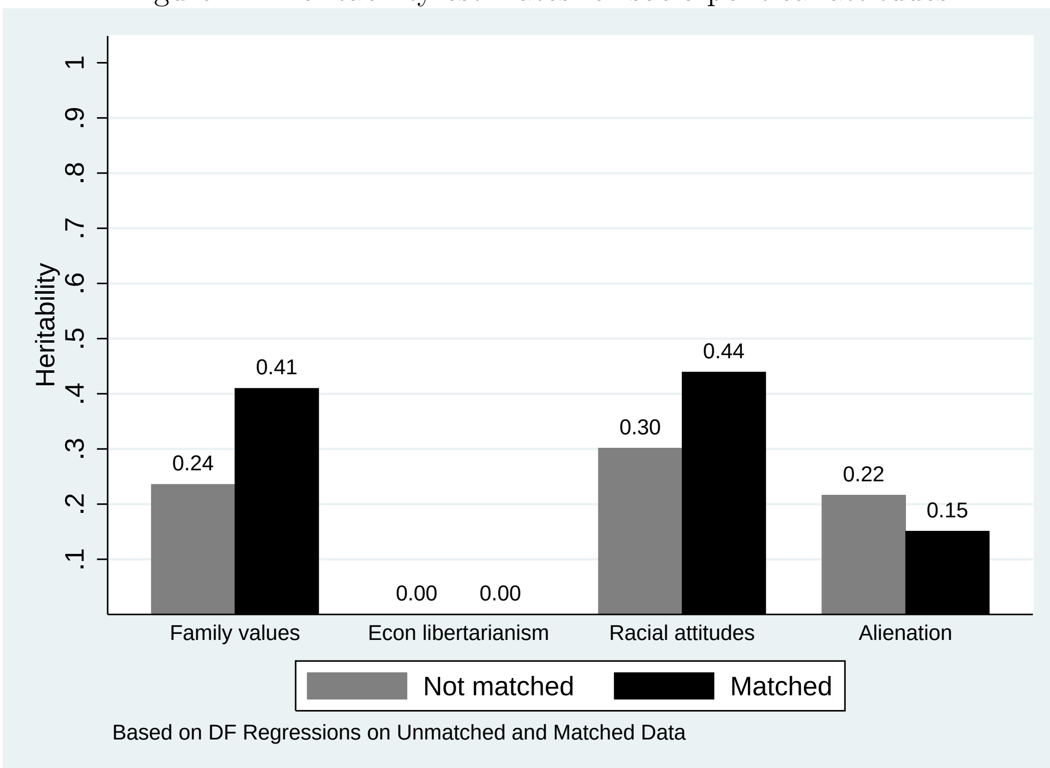


Figure 43: Heritability estimates for measures of altruism

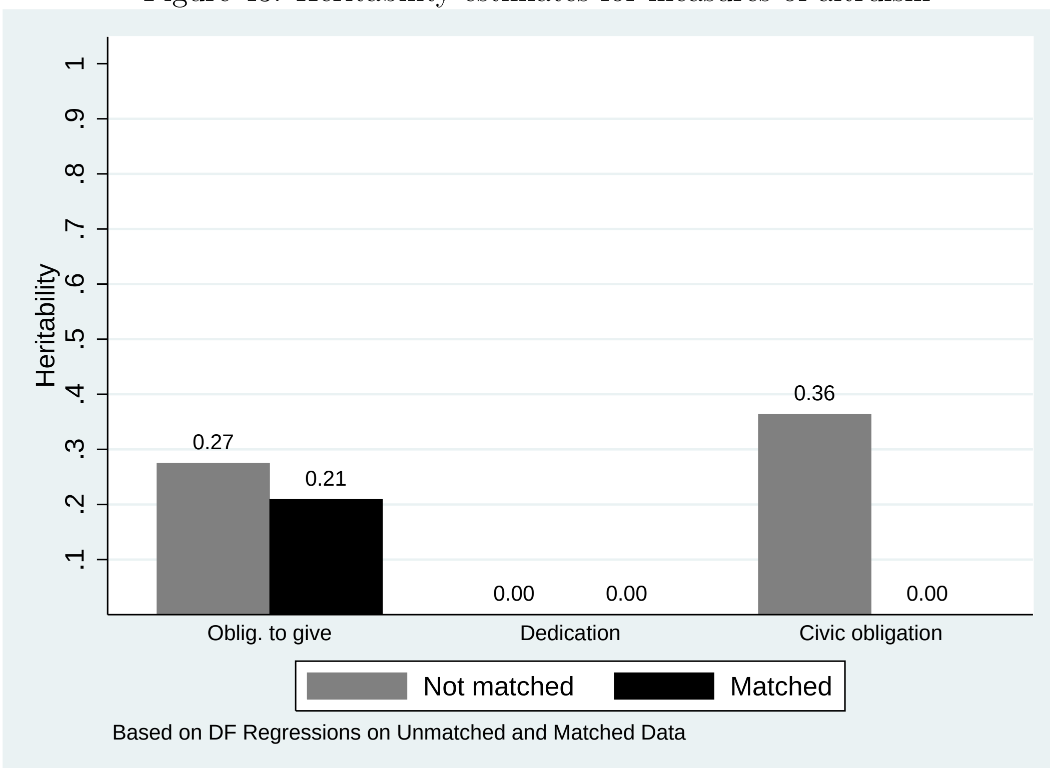
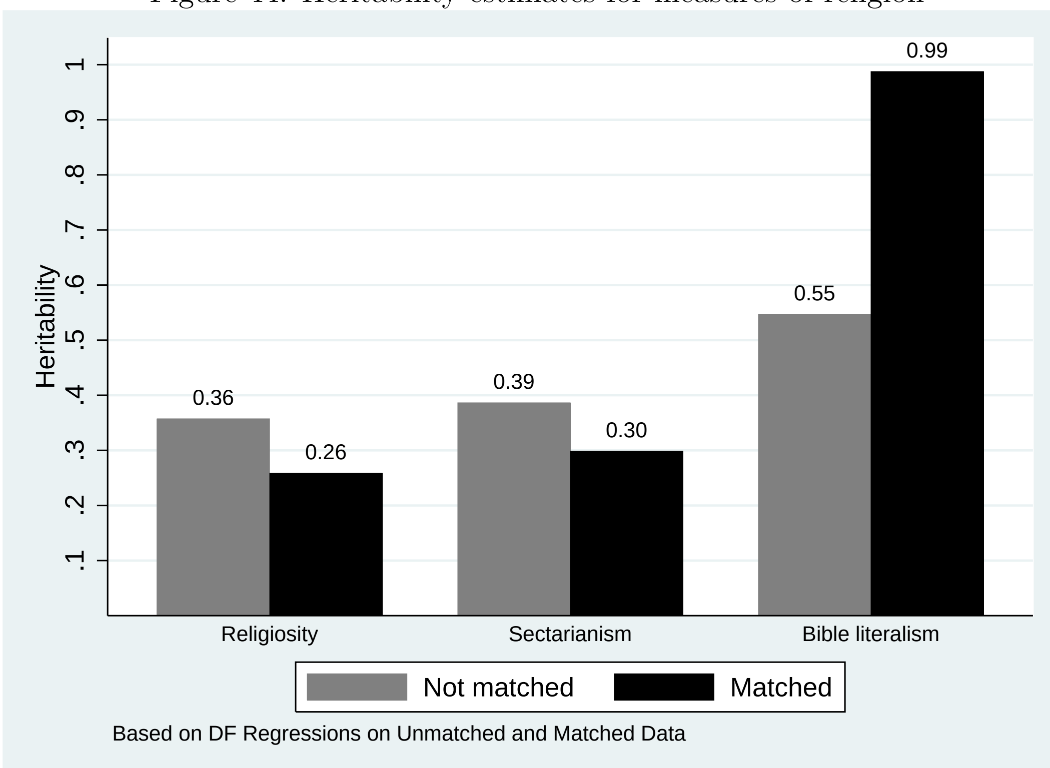


Figure 44: Heritability estimates for measures of religion



6 Conclusion

In the introduction, I argued that it was useful to evaluate classic twin studies in terms of how well they isolate genetic causes. Toward this end, I considered the strengths and weaknesses of twin studies in terms of how well these studies approximate a true experiment. I concluded that, insofar as zygosity (MZ & DZ) is randomly distributed across the population, the twin study design approximates a true experiment quite well. Although DZ twinning is not randomly distributed across the population, the patterning of DZ twin births does not appear to be related to the extent of similarity between DZ twins. Meanwhile, MZ twinning appears to occur at random. Since the mechanism that determines zygosity appears to be unrelated to any confounding variables, twin studies seem to have high internal validity.

As I argued previously, the chief doubts about twin studies are related to their external validity. We can be confident that twin studies reveal the extent to which variation in *twin* behavior can be attributed to variation in genetics. But there is uncertainty about the extent to which the results of twin studies can be extrapolated to the non-twin population. The underlying cause of MZ similarity is certainly the fact that they are genetic clones. But MZ similarity may be at least partly a function of the unique environment that results from genetic clones growing up together. In other words, genetic similarity may be a necessary but insufficient cause of the similarity in outcomes observed between MZ twins. Environmental similarity may play a key role in mediating the relationship between similarity in genes and similarity in outcomes between co-twins. If so, the equal environments assumption is invalid, and the estimates of genetic influence from twin studies may not apply to the population at large. The goal of this study was to evaluate the equal environments assumption – the Achilles Heel of the classic twin study.

After concluding that the equal environments assumption was the main weakness of twin studies, I reviewed the literature that tested this assumption. Soon after the first twin studies were conducted, researchers recognized that the environments of MZ twins were more similar than the environments of DZ twins. The question was whether the effects of zygosity on outcomes were confounded with the effects of environmental similarity.

Authors of early studies testing the EEA generally concluded that the data provided unequivocal support for the EEA. My review of this literature showed that the evidence was more ambiguous. For example, Scarr (1968) concluded that her data supported the EEA, but my interpretation of her data suggests that a more nuanced conclusion may be warranted. Her data suggested that genetic similarity matters more than environmental similarity, but environmental similarity still seems to have an effect. In other words, the upward bias affecting heritability estimates in twin studies may be modest but nonzero.

My reanalysis of the data used in the first extensive examination of the EEA by Loehlin and Nichols (1976) also revealed problems with the equal environments assumption. Loehlin and Nichols (1976) found few significant correlations between parents' ratings of twins' environments and twins' characteristics, and concluded that the EEA was supported. According to Joseph (2004), the use of parents' ratings of environmental similarity is problematic because parents are likely to minimize the ways in which twins are treated differently. This problem can be avoided by using twins' own ratings of environmental similarly. The correlations between twins' own ratings of environmental similarity and absolute differences in twin outcomes were generally low in magnitude but some correlations were significant. My reanalysis of the Loehlin and Nichols (1976)'s data revealed modest violations of the EEA. I estimated the heritability of a variety of traits in L&N's data – both with and without adjusting for differences in environmental similarity between MZ and DZ twins. On average, adjusting for environmental similarity reduced estimates of heritability by about 34%.

Reexamining the evidence presented in another frequently cited study, Scarr and Carter-Saltzman (1979), I again found that the EEA had not been thoroughly vetted. Scarr and Carter-Saltzman (1979) regressed absolute differences between co-twins' cognitive test scores on zygosity and on measures of perceived differences between twins. The authors concluded that the EEA was valid because the effects of perceived differences between twins were insignificant when zygosity was controlled. But the authors included five measures of perceived similarity in a regression on data with 226 cases. Since the five measures of perceived similarity are likely to be highly intercorrelated, one is unlikely to be able to find a significant effect of any of the variables when they are included together in the same regression. Like previous studies, Scarr and Carter-Saltzman (1979) did not give the EEA enough chance to fail.

The evidence presented in favor of the EEA is stronger and more consistent in recent studies than the evidence from the earlier period. Collectively, recent studies tested a much broader array of measures of environmental similarity than previous studies. Very few violations of the equal environments assumption were found. However, the results of recent studies are limited in scope. Most studies focused exclusively on psychological characteristics such as personality traits and mental disorders. The limited scope of recent tests of EEA is particularly problematic now that twin studies are increasingly applied to outcomes of sociological interest, such as political attitudes and educational attainment.

In my analysis, I calculated estimates of heritability with and without adjusting for a variety of measures of environmental similarity. My study included a greater variety of measures of environmental similarity than previous research. Included in these controls is a measure of intimacy, which was shown to confound the effects of genetic effects by Penninkilampi-Kerola and Moilanen (2005). My study also tested the EEA using a larger and more diverse set of outcomes than any previous study. Another innovation of the present study is the application of propensity score matching to the twin study method. Unlike regression, the matching technique elucidates the fundamental comparison that is being made between MZ and DZ twins matched on environmental similarity.

Overall, the results of my analysis indicate that violations of EEA are both larger and more numerous than previous research suggests. I summarize the effects of adjusting for environmental similarity on heritability estimates in Figure 45. I omit outcomes for which unadjusted estimates of heritability were zero. Of the heritability estimates for 47 outcomes summarized in Figure 45, 16 were reduced by more than 25%. Heritability estimates for 17 outcomes were reduced modestly (between 0 and 25%) when matching on environmental similarity. In general, adjustments for environmental similarity led to weak-to-moderate reductions in most estimates of heritability.

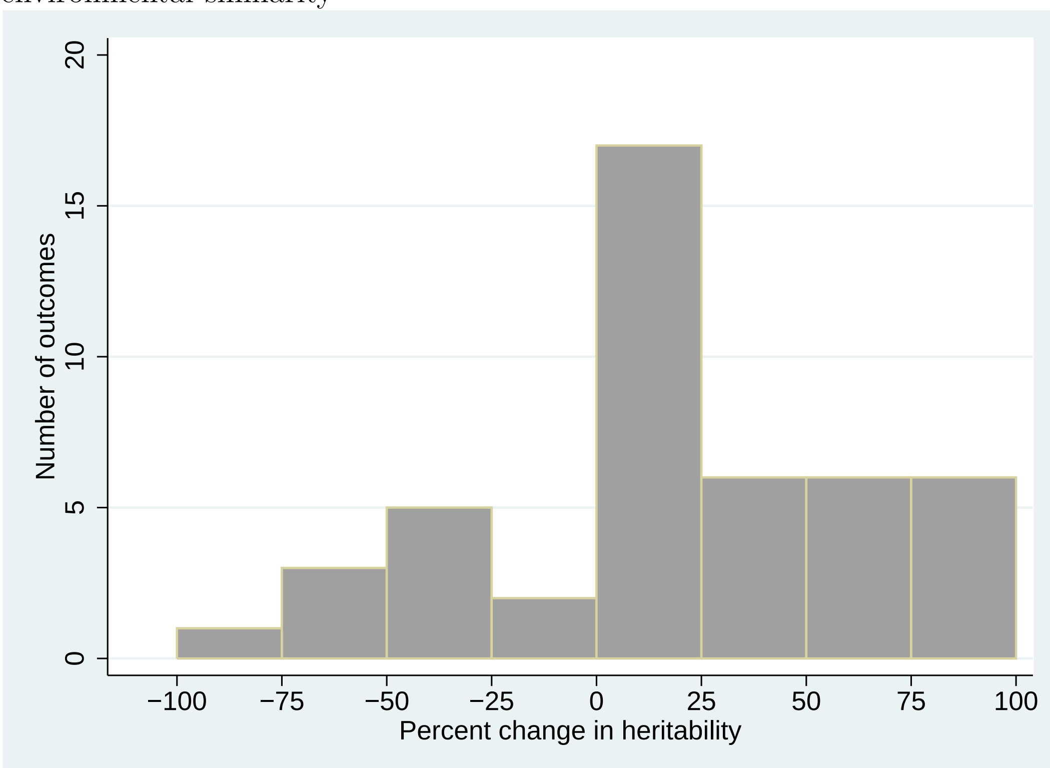
Reductions in heritability estimates are not patterned in any discernable way. Estimates of heritability for sociological outcomes are no more likely to be reduced than are estimates of heritability for psychological outcomes.

It is likely that the EEA did not hold up as well in this study as it did in some previous studies because the present study incorporated more control variables.

The present study has provided the most in-depth synthesis of findings related to the EEA and the most thorough analysis of EEA. The general conclusion is that the equal environments assumption holds up in some cases but not in others. It is likely that estimates of heritability for most outcomes are biased upward at least modestly. Since it is difficult to predict when bias will be substantial, all future twin studies should incorporate extensive measures of environmental similarity. Adjusting for the bias induced by environmental similarity is likely to reduce but not eliminate the apparent effects of genes.

Thus far, sociologists who have engaged with evidence from twin studies have often taken one of two tacks: (1) accept the methodology and conclusions as essentially valid (e.g. Jeremy Freese, Jason Schnittker) or (2) suggest broad-based skepticism from a theoretical perspective (e.g. Andrew Perrin). The methodological critique of the twin method presented here supports a middle ground between complete acceptance and complete rejection of the results of twin studies.

Figure 45: Percentage change in heritability of outcomes after matching on measures of environmental similarity



References

Alford, J., Funk, C., and Hibbing, J. 2005. Are political orientations genetically transmitted? *American Political Science Review* 99:153–167.

Alwin, D. 2007. Margins of Error: A Study of Reliability in Survey Measurement. Wiley, Hoboken, NJ.

Association, A. P. 2000. Diagnostic and Statistical Manual of Mental Disorders. American Psychiatric Publishing, Arlington, VA.

Bearman, P. and Bruckner, H. 2002. Opposite sex twins and adolescent same-sex attraction. *American Journal of Sociology* 107:1179–1205.

Borkenau, P., Riemann, R., Angleitner, A., and 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., and Tellegen, A. 1990. Sources of human psychological differences: the Minnesota study of twins reared apart. *Science* 250:223–228.

Brim, O., Baltes, P., Bumpass, L., Cleary, P., Featherman, D., Hazzard, W., Kessler, R., Lachman, M., Markus, H., Marmot, M., Rossi, A., Ryff, C., and Shweder, R. 2008. National survey of midlife development in the United States 1995-1996. Downloaded from www.icpsr.umich.edu.

Clogg, C., Petkova, E., and Haritou, A. 1995. Statistical methods for comparing regression coefficients between models. *American Journal of Sociology* 100:1261–1293.

Cronk, N. J., Slutske, W. S., Madden, P. A., Bucholz, K., Reich, W., and 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.

DeFries, J. and Fulker, D. 1985. Multiple regression analysis of twin data. *Behavior Genetics* 15:467–473.

DERKS, E. M., DOLAN, C. V., AND BOOMSMA, D. I. 2005. A test of the equal environment assumption (EEA) in multivariate twin studies. *Twin Research and Human Genetics* 9:403–411.

DION, K., BERSCHEID, E., AND WALSTER, E. 1972. What is beautiful is good. *Journal of Personality and Social Psychology* 24:285–290.

EAVES, L., HATEMI, P., PROM-WOMLEY, E., AND MURRELLE, L. 2008. Social and genetic influences on adolescent religious attitudes and practices. *Social Forces* 86:1621–1646.

ERIKSSON, M., RASMUSSEN, F., AND TYNELIUS, P. 2006. Genetic factors in physical activity and the equal environment assumption – the Swedish young male twins study. *Behavior Genetics* 36:238–247.

GOULD, W., PITBLADO, J., AND SRIBNEY, W. 2003. Maximum Likelihood Estimation with Stata. Stata Press, College Station, TX.

GUNDERSON, E., AI-LIN, T., SELBY, J., CAAN, B., MAYER-DAVIS, E., AND 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.

HARRIS, J. 1999. The Nurture Assumption. The Free Press, New York, NY.

HETTEMA, J. M., NEALE, M. C., AND 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., AND 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.

JENSEN, A. 1998. The G Factor: The Science of Mental Ability. Praeger, Westport CT.

JONES, H. E. AND WILSON, P. T. 1932. Reputation differences in like-sex twins. *Journal of Experimental Education* 1:86–92.

Joseph, J. 2004. The Gene Illusion: Genetic Research in Psychiatry and Psychology Under the Microscope. Algora Publishing, New York, NY.

Kasriel, J. and Eaves, L. 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. and 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., Neale, M. C., Kessler, R., and Heath, A. 1993. A test of the equal-environment assumption in twin studies of psychiatric illness. *Behavior Genetics* 23:21–27.

King, G., Keohane, R. O., and Verba, S. 1994. Designing Social Inquiry: Inference in Qualitative Research. Princeton University Press, Princeton, NJ.

Klump, K. L., Holly, A., Iacono, W. G., McGue, M., and Wilson, L. E. 2000. Physical similarity and twin resemblance for eating attitudes and behaviors: a test of the equal environemnts assumption. *Behavior Genetics* 30:51–58.

Knoke, D. and Burke, P. 1980. Log-Linear Models. Sage Publications, Newbury Park, CA.

Kohler, H.-P. and Rodgers, J. 1999. DF-like analyses of binary, ordered, and censored variables using probit and tobit approaches. *Behavior Genetics* 29:221–232.

Kohler, H.-P. and Rodgers, J. 2000a. DF-analyses of heritability with double-entry twin data: Asymptotic standard errors and efficient estimation. *Behavior Genetics* 31:179–191.

Kohler, H.-P. and Rodgers, J. 2000b. Programs for DF-like analyses of binary, ordered and censored variables using probit and tobit approaches. `http://www.ssc.upenn.edu/~hpkohler/data-and-programs/twbioprobits/twbioprobits.html`.

Levene, H. 1960. Robust tests for equality of variances. *In* I. Olkin (ed.), Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling. Stanford University Press, Stanford, CA.

Loehlin, J. and Nichols, R. 1976. Heredity, Environment and Personality: A Study of 850 Sets of Twins. University of Texas Press, Austin.

Lytton, H. 1977. Do parents create, or respond to, differences in twins? *Developmental Psychology* 13:456–459.

Martin, M. 2008. The intergenerational correlation in weight: How genetic resembalance reveals the social role of families. *American Journal of Sociology* 114:S67–S105.

Matheny, A. P., Wilson, R. S., and Dolan, A. B. 1976. Relations between twins' similarity of appearance and behavioral similarity: testing an assumption. *Behavior Genetics* 6:343–351.

McCrae, R. and John, O. 1992. An introduction to the five factor model and its applications. *Journal of Personality* 60:175–215.

Morgan, S. and Winship, C. 2007. Counterfactuals and Causal Inference: Methods and Principles for Social Research. Cambridge University Press, New York, New York.

Morris-Yates, A., Andrews, G., Howie, P., and Henderson, S. 1990. Twins: a test of the equal environments assumption. *Acta Psychiatry Scandinavia* 81:322–326.

Mowrer, E. R. 1954. Some factors in the affectional adjustment of twins. *American Sociological Review* 19:468–471.

Nielsen, F. 2006. Achievement and ascription in educational attainment: Genetic and environmental influences on adolescent schooling. *Social Forces* 85:193–216.

Pedersen, N., Plomin, R., Nesselroade, J., and McClearn, G. 1992. A quantitative genetic analysis of cognitive abilities during the second half of the life span. *Psychological Science* 3:346–353.

Penninkilampi-Kerola, V. and Moilanen, I. 2005. Co-twin dependence, social interactions, and academic achievement: a population-based study. *Journal of Social and Personal Relationships* 22:519–541.

Powers, D. and Xie, Y. 2000. Statistical Methods for Categorical Data Analysis. Academic Press, San Diego, CA.

Price, B. 1950. Primary biases in twin studies: a review of prenatal and natal differences producing factors in monozygotic pairs. *American Journal of Human Genetics* 2:293–352.

Rodgers, J., Kohler, H.-P., McGue, M., Behrman, J., Petersen, I., Bingley, P., and Christensen, K. 2008. Education and cognitive abiilty as direct, mediated, or spurious influences on female age at first birth: behavioral genetic models fit to Danish twin data. *American Journal of Sociology* 114:S202–S232.

Rose, R. J., Koskenvuo, M., Kaprio, J., Seppo, S., and 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. 1994. The Limits of Family Influence. The Guilford Press, New York, NY.

Scarr, S. 1968. Environmental bias in twin studies. *Eugenics Quarterly* 15:34–50.

Scarr, S. and Carter-Saltzman, L. 1979. Twin method: Defense of a critical assumption. *Behavior Genetics* 9:527–542.

Smith, R. T. 1965. A comparison of socioenvironmental factors in monozygotic and dizygotic twins, testing an assumption. *In* S. Vandenberg (ed.), Methods and Goals in Human Behavior Genetics. Academic Press, New York.

Stack, S. 1992. The effect of country music on suicide. *Social Forces* 71:211–219.

Stocks, P. 1930. A biometric investigation of twins and their brothers and sisters. *Annals of Eugenics* 4:49–108.

Vandenberg, S. 1966. Contributions of twin research to psychology. *Psychological Bulletin* 66:327–352.

Watson, J. and Crick, F. 1953. A structure for deoxyribose nucleic acid. *Nature* 171:737–738.

Wilson, P. T. 1934. A study of twins with special reference to heredity as a factor determining differences in environment. *Human Biology* 6:324–354.

Vita

Jacob Felson

William Paterson University
Department of Sociology
300 Pompton Road
Wayne, NJ 07470

Phone: (814) 441-0696
Email: felsonj@wpunj.edu

Education

B.A. Sociology, University of Chicago, 2001.

M.A. Sociology, Pennsylvania State University, 2004.

Ph.D. Sociology, Pennsylvania State University, 2009.

Employment

Research assistant, Department of Sociology, Pennsylvania State University, 2001-2007.

Instructor, William Paterson University, 2007–2009.

Refereed publications

Amy Adamczyk and Jacob Felson. Fetal Positions: Religion and Pregnancy Resolution, 2008, *Social Science Quarterly* 89, 17–38.

Jacob Felson and Heather Kindell. The Elusive Link between Conservative Protestantism and Conservative Economics, 2007, *Social Science Research* 36, 673–687.

Duane Alwin, Jacob Felson, Edward Walker and Paula Tufis. Measuring Religious Identities, 2006, *Public Opinion Quarterly 70: 530–564.*

Amy Adamczyk and Jacob Felson. Friends' Religiosity and First Sex, 2006, *Social Science Research* 35, 924–947. *Received student paper award, ASA section, Children and Youth