4-2010

The Genetic Heritability of Survey Response Styles

Levente Littvay
University of Nebraska-Lincoln, levi@huskers.unl.edu

THE GENETIC HERITABILITY OF SURVEY RESPONSE STYLES

by

Levente Littvay

A THESIS

Presented to the Faculty of

The Graduate College at the University of Nebraska

In Partial Fulfillment of Requirements

For the Degree of Master of Science

Major: Survey Research and Methodology

Under the Supervision of Professor Robert F. Belli

Lincoln, Nebraska

May, 2010

THE GENETIC HERITABILITY OF SURVEY RESPONSE STYLES

Levente Littvay, M.S.

University of Nebraska, 2010

Adviser: Robert F. Belli

This study assesses the genetic heritability of various survey response styles using a classical twin design. The National Survey of Midlife Development in the United States (MIDUS) collected in 1995-96 included an oversample of twins with self-reported zygosity along with a large number of survey items that allowed for the assessment of acquiescent and extreme response style. The MIDUS singleton sample was used for the careful development of appropriate and reliable measures of these traits. The second wave of the MIDUS (2005-06), was used to assess the sources of survey response trait stability. Acquiescence appears to have a sizable and significant heritability component (~25% to 34%) and is not influenced by the environment co-twins share; extreme response has a smaller and nonsignificant heritable component (~20%) and a more sizable shared environmental effect (~32%). Additive genetic and common environmental effects drive the test-retest stability of traits. The variation explained by these two effects also wholly contributes to the covariation across waves as well. Controlling for income and education does not change results. Results suggest it is inappropriate to treat acquiescence and extreme response as a single survey response style phenomena and call for the exploration of specific genes in the case of acquiescence and social and familial environmental predictors of extreme response.

Acknowledgements:

Primarily, I would like to thank my committee. Without the infinitely helpful guidance of Bob Belli and John Hibbing this project could never have become what it is now. I would like to thank people I received comments from, especially participants of The 2009 International Workshop on Statistical Genetics and Methodology of Twin and Family Studies, notably Lindon Eaves, Nick Martin, Mike Neale, Pete Hatemi and Stacey Cherny, who all commented on a poster version of the first iteration of this study. Thanks to the participants of the Doctoral Seminar at the Central European University, Department of Political Science where I had the pleasure and opportunity to present this thesis, with special thanks to my discussant Gabor Toka. Many thanks to those who commented on various versions: Zoltan Fazekas, Matthew Hibbing and Brennen Bearnes. Megan Thornton, my wife, deserves a special big thank you for all the support, proofreading and language editing. Many, many thanks to Barb Rolfes for all the logistics help. To everyone at SRAM, it becomes more and more apparent over time how much I learned from you, thank you. A very special thanks goes to Allan McCutcheon for recruiting me into the Survey Research and Methodology program, offering his guidance during my time there, and introducing me to Tamas Rudas though whom I received my dream job straight out of college. I feel a small hole in my life for not being able to work with him on this final project. This thesis is dedicated to Dindi who left us in September 2009 a day short of her 101st birthday. Her birth predates modern survey research or political science research. She will forever be missed.

Introduction

Survey researchers have long studied questions such as why do some people acquiesce on questionnaires without regard to the specific questions asked (Cheung and Resvold 2000, Javeline 1999, Kuethe 1960, Lentz 1938, Marin et al 1992, Ross and Mirowsky 1996, Smith 2004)? Why do some people select extreme responses while others deviate more to the center values (Arce-Ferrer 2006, Berg and Colier 1953, Cheung and Resvold 2000, Chun et al 1974, Clarke 2000, Clarke 2001, de Jong et al 2008, Greenleaf 1992, Hui and Triandis 1989, Lewis and Taylor 1955, Light and Zax 1965, Marin et al 1992)? What predicts satisficing (Belli et al. 1999, Krosnick 1991, Holbrook et al 2003)? These issues have been studied since researchers first suspected that surveys are not entirely error free measurement instruments (Lorge 1937, Lentz 1938). They have been called survey response styles (Bachman and O'Malley 1984, Baumgartner and Steenkamp 2001, Chen et al 1995, Gilman et al 2008, Jackson and Pacine 1961, Johnson et al 2005, Lamm and Keller 2007, Stening and Everett 1984, van Herk et al 2004), response sets (Berg and Colier 1953, Cheung and Resvold 2000, Cronbach 1946, Fricke1956, Hui and Triandis 1985, Kuethe 1960, Landsberger Saavedra 1967), response effects (Ayidiya and Mcclendon 1990, Javeline 1999, Kieslerey and Sproull1986, Narayan and Krosnick 1996) and response bias (Carr and Krause 1978, Holbrook et al 2003, Smith 2004). Two constructs that are often examined are acquiescence and extreme response (Arce-Ferrer et al 2006, Ayidiya and McClendon 1990, Baumgartner and Steenkamp 2001, Berg and Collier 1953, Biemer 2001, Carr and Kraus 1978, Chen et al 1995, Cheung and Rensvold 2000, Chun et al 1974, Clarke 2000, Clarke 2001, de Jong et al 2008, Fricke 1956, Gilman et al 2008, Holbrook et al 2003, Hui and Triandis 1985, Jackson and Pacine 1961, Javeline 1999, Johnson et al 2005, Jordan et al 1980, Kiesler and Sproull 1986, Kuethe 1960, Lamm and Keller 2007, Landsberger and Saavedra 1967, Lenski and Leggett 1960, Lewis and Taylor 1955, Light and Zax 1965, Marin et al 1992, Narayan and Krosnick 1996, Ross and Mirowsky 1996, Roster et al 2007, Shulman 1973, Smith 2004, Stening and Everett 1984, van Herk et al 2004, Watkins and Cheung 1995).

A large number of environmental independent variables have been explored and identified as predictors of acquiescence and extreme response[1]. It is also well established that acquiescence and extreme response are stable traits in test-retest situations even when several years pass between tests (Berg and Collier 1953, Bachman and O'Malley 1984). Theoretical and empirical research tied survey response styles to cognitive ability, altruism and accountability (Krosnick 1991, Krosnick 1999, Belli et al 1999), which are either directly or through personality traits could provide a mechanism for survey response styles to be genetically heritable (for studies on the heritability of personality see Eysenck and Prell 1951, Jang et al 1998, Loehlin and Nichols 1976, Rose et al 1988, Rushton et al 1986, Bouchard et al 1990, Pederson et al 1988, Horn et al 1976, Tellegen et al 1988, Heath et al 1992, McCrae and Costa 2003, Riemann et al 1997). Based on this information it is a serious plausibility that people’s genetic makeup could influence

[1] For a detailed review see next section.

how they respond to surveys. This thesis tests the proposition that survey response styles are genetically heritable.

Behavior genetics has been a prevalent subfield in psychology (Jang et al 1998, van Beijsterveldt and van Baal 2002) but other social sciences have been slow to consider the possibility that genes can influence behaviors. Alford, Funk and Hibbing's (2005) *American Political Science Review* (*APSR*) article exploring the genetic transmission of political attitudes laid the groundwork for this type of research within political science. Findings of this article have appeared in popular media (such as The New York Times) and the article became the most downloaded article of the *APSR*. It was cited by the editor as the most important article published by the journal in its 100 year history (Sigelman 2006). Since then, several pieces exploring how genetics could influence political behavior have appeared in top political science journals such as *Political Analysis* (Medland and Hatemi 2009), The *Journal of Politics* (Hatemi et al 2009, Hatemi et al 2009, Fowler and Dawes 2008, Dawes and Fowler 2009) and *Political Research Quarterly* (Settle et al 2009, Hatemi et al 2009). Two articles on the topic are also published in *Science* (Oxley et al 2008, Fowler and Schreiber 2008).

Economists are also starting to publish work on genetic predictors of economic behavior in prestigious journals such as the *Proceedings of the National Academy of Science* (Cesarini et al 2008) and *The Quarterly Journal of Economics* (Cesarini et al 2009). The topic has also gained press within sociology (Guo and Stearns 2002, Eaves et al 2008, also see special issue on biological predictors of social behavior in *Social Forces* September 2006, Issue 85:1). Behavior geneticists have also developed an interest in the “phenotypes” studied by social scientists. The Behavior Genetics Association’s journal, *Behavior Genetics*, has already published articles on the topic of politics (Hatemi et al 2007, Eaves and Hatemi 2008) and also genes’ interactions with socio-economic variables (Harden et al 2007).

While the use of biological predictors of political behavior are not new to the field (Madsen 1985, Madsen 1986) recent publications have spawned debates within political science (Charney 2008a, Alford et al 2008a, Hannagan and Hatemi 2008 and Charney 2008b, Beckwith and Morris 2008, Alford et al 2008b). Similar debates have taken place in other social science disciplines such as Sociology (Horwitz et al 2003b, Freese and Powel 2003, Horwitz et al 2003 a), Educational Psychology (Richardson and Norgate 2005), and Psychology (Joseph 2001). After a review of the debates, this thesis brings our own field of survey research on board with the behavior genetic research agenda through assessing the heritability of survey response styles.

Survey Response Sets and Survey Response Styles

Survey response styles are a problem for the measurement of unbiased true scores of any latent construct. While random measurement error is manageable by statistical methods and, even if unmanaged, only leads to type II error, systematic bias to survey response will distort the measures of any construct of interest. Any such definable and identifiable phenomena that distort survey measurement are a concern for survey researchers. It is no surprise that survey response styles received attention as early as 1937 when Lorge identified a “halo” that pollutes attitudinal responses in a consistent direction independent of question coding (1937). In 1938 Lentz specifically used the term acquiescence in describing a “factor in the measurement of personality” (1938). Cronbach coined the term *response sets* as the “tendency causing a person consistently to make different responses to test items than he would have made had the same content been presented in a different form” (1946). The idea of response sets in the broader survey measurement field gained leverage in the 1950’s (Berg and Collier 1953, Lewis and Taylor 1955, Fricke 1956, Kuethe 1960).

Since then extreme response and acquiescence has often been considered together under the same heading of response set, response effect, response style or response bias (Shulman 1973, Ayidiya and McClendon 1990, Hui and Triandis 1985, Gilman et al 2008, Baumgartner and Steenkamp 2001, Cheung and Rensvold 2000). There have been attempts to nuance the distinctions between the terms. For example, Watkins and Cheung (1995) referred to systematic distortion in one direction independent of item content as a response style; response set was defined as conscious or unconscious desire to project a particular self-picture by the respondent (see also Lanyon 1982). Such nuances did not take hold in the literature and are rarely acknowledged as both phenomena manifest in the same indicator of a certain style of response. Their separation is impossible with most tools used by researchers studying the phenomena.

Predictors of Survey Response Styles

As early as 1960 researchers had started to explore the predictors of response sets. The most commonly used predictor is education (or social class operationalized by a measure of education) and is consistently found to be a negative predictor of acquiescence and extreme response (Lenski and Leggett 1960, Landsberger and Saavedra 1967, De Jong et al 2008, Stening and Everett 1984, Narayan and Krosnick 1996, Marin et al 1992, Carr and Kraus 1978). Along the same lines Jackson and Pacine (1961) connected response sets to academic achievement, Light and Zax (1965) to intelligence and Belli et al (1999) connected a related construct, satisficing response, to cognitive ability (Krosnick 1991).

Other studies of response styles include cross-cultural exploration (Javeline 1999, Watkins and Cheung 1995, Gilman et al 2008, Clarke 2000, Clarke 2001, Chun et al 1974, Hui and Triandis 1989, Marin et al 1992, Cheung and Rensvold 2000, van Herk et al 2004, Johnson et al 2005, Lamm and Keller 2007, Arce-Ferrer et al 2006, Smith 2004, Roster et al 2006, Stening and Everett 1984, Baumgartner and Steenkamp 2001, De Jong et al 2008, Chen et al 1995, Ross and Mirowsky 1996) and mode effects (Holbrook et al 2003, Jordan et al 1980, Kiesler and Sproull 1986, Biemer 2001). Because the data used in this thesis is collected in the United States through a self-administered questionnaire, the results presented hold mode and cultural factors constant. Future research could address other modes and contexts.

Satisficing and Survey Response Styles

Krosnick's seminal article in *Applied Cognitive Psychology* introduced and discussed hypothetical sources of satisficing in surveys (Krosnick 1991). Acquiescence (or "agreeing with assertions") is explicitly listed as a form of satisficing but extreme response is not (Krosnick 1991:218). It could be argued that extreme response could fit under the category "non-differentiation in using rating scales" as extreme response style means that the respondent is not fully utilizing the entire width of the scale and therefore not differentiating between stronger and less strong response categories (Krosnick 1991:218). But Krosnick thought of this more as differentiating across items and not within (Krosnick 1991). In other words non-differentiation applies to the respondent giving the same response to a large number of questions, and not providing different end point responses as would happen with respondents with extreme response styles.

Krosnick (1991) identified three sources of satisficing behavior: (1) inherent difficulty of the task, (2) respondent's ability to perform the required task and (3) respondent's motivation to perform the task. Task difficulty is entirely dependent on the questions. So if acquiescence or extreme response (or any other satisficing tendency) is measured using a fixed set of items this source of the behavior cannot be considered as a cause for variation across respondents.

Respondent's ability, as already discussed in the previous section, has also been referred to as cognitive ability. When direct measures of cognitive ability were not available education was often used as a predictor of response styles. While

Krosnick quickly dismisses that IQ, a highly heritable construct (Devlin et al 1997, Visscher et al 2006), is not the same as cognitive ability, studies that directly looked at cognitive ability also found the trait to be highly heritable (Plomin et al 1994, Rietveld et al 2003). This suggests a possible mechanism through which satisificng survey response could be genetically heritable. Also, familiarity with the topic and the availability of preconditioned attitudes on the topic are cited as influencing respondent ability and motivation (discussed more in detail in the next paragraph). Given the broad range of topics in the survey used in this thesis, it is reasonable to assume that, on average, no certain topic will influence any individual respondent's average survey response style scores.

Finally respondent motivation is broken down into the respondent's need to engage in high levels of cognition (Krosnick 1991). Ability to perform high-level cognition is not the same as having the will to perform high-level cognition. Performance of high-level cognition will be dependent on the individual's intrinsic payoff for performing the mental exercises. Such payoffs could be related to personality, which is known to be highly heritable (Eysenck and Prell 1951, Jang et al 1998, Loehlin and Nichols 1976, Rose et al 1988, Rushton et al 1986, Bouchard et al 1990, Pederson et al 1988, Horn et al 1976, Tellegen et al 1988, Heath et al 1992, McCrae and Costa 2003, Riemann et al 1997). Secondly, motivation is also related to how important and/or useful the respondent sees the survey to be. While this could be influenced by the interviewer and other communications from the side of researchers, but given a fixed contact procedure this influence should be relatively constant in a specific survey effort such as the one used in this study. On the other hand, individual's level of altruism could influence their views on the importance of the survey. Agent based simulation models have identified mechanisms through which altruistic behavior could evolve (Hammond and Axelrod 2006, Nemeth and Takacs 2007). Assuming that these simplistic evolutionary simulations translate into the real world, survey satisficing survey behavior could be genetically heritable through the heritability of altruistic tendencies. And this appears to be the case as questionnaire based empirical research using a twin design yielded 56% heritability of altruism (Rushton et al 1986). Finally, respondent motivation has been linked to accountability. If a respondent feels that they are accountable for their answers, that they might have to justify them in the future, that could lead to less survey satisficing behavior. While the survey protocol could certainly influence the feeling of accountability in a respondent, it could also be influenced by the respondent's personality traits such as conscientiousness and possibly even neuroticism if the respondent worries about being accountable. Tupes and Christal (1958, 1961) identified five components that explain the majority of variation in someone's personality. These five components were later relabeled as the big 5. Conscientiousness and neuroticism are two of the (big) five personality constructs along with agreeableness, extraversion and openness. All of the big 5 personality traits are known to be about 50% heritable (Jang et al 1998), through these traits survey satisificing behavior could also be influenced by a respondent's genetic dispositions. And, finally, the length of the interview influences motivation but in the survey analyzed here, the length was more or less constant given the small number of skip patterns.

In addition to his research on satisficing survey response Krosnick also directly addressed possible sources of survey acquiescence in a later study (Krosnick 1999). Krosnick divided explanations of satisficing into sociological and psychological factors. The sociological explanation states that if the "researchers and interviewers are perceived as being of higher social status the respondent may defer to them out of courtesy". From this perspective the respondent's status and self-efficacy could both negatively influence acquiescing behavior. The psychological explanation states that a person's agreeableness will determine how much they acquiesce on a survey. Agreeableness is a Big 5 personality trait known to be highly heritable (Jang et al 1998) and Krosnick specifically states that because of its links to agreeableness, acquiescence "may have genetic roots" (Krosnick 1999:553).

Expectations

The early literature treated survey response as a personality trait (Lentz 1938) but this presumption has not been followed up as the personality literature has been refined. But today we know that personality is highly heritable and for this reason it is reasonable to presume that survey response styles are also heritable.

The satisficing literature only specifically encompasses acquiescence though extreme response could also implicitly fit satisficing survey behavior. Krosnick has identified three mechanism though which, I posit, survey satisficing behavior could have genetic roots. It is cognitive ability, altruism and accountability, with the latter being closely related to highly heritable Big 5 personality traits: conscientiousness and neuroticism (Krosnick 1991). Additionally Krosnick's psychological explanation of acquiescence ties it to the third highly heritable Big 5 personality trait: agreeableness (1999).

It is for these reasons I expect that both survey acquiescence and extreme response style is genetically heritable. The theoretical evidence for suspecting heritability of survey acquiescence is stronger then it is for extreme response style.

Classical Twin Design to Estimate Heritability

The classical twin design uses phenotypic data from twins reared together to estimate the relative impact of heritability, common and unique environmental factors. We know that identical twins share 100% of their genome. We also know that, on average, fraternal twins share 50% of their genome. Given the availability of data on twin pairs, it is possible to estimate how much of any trait is influenced by the twins' genome (most often abbreviated as A), the environment they share (also referred to as common environment and abbreviated as C), or the environment they do not share (or unique environment abbreviated as E). The model to estimate these is often called the ACE model. For example, if monozygotic (or MZ) twin pairs correlate on the trait more than dizygotic (or DZ) twin pairs, we can infer that genes influence the trait, whereas if identical twins and fraternal twins correlate equally, there is a good chance their shared environment influences the phenotype. Additional deviations from perfect correlations are attributed to the unique environmental influences. This unique environmental component also includes measurement error as it is another cause of downward deviation from perfect co-twin correlation. In other words the model, to produce unbiased results, assumes no measurement error. The proportion of error cannot be separated from true environmental influences therefore special emphasis should always be placed on careful phenotypic measurement with minimal measurement error when using a twin design.

Until the 1970's, the classical twin design used no more than correlations to estimate the percent of genetic and environmental influences (Falconer 1960, Alford et. al. 2005, Medland and Hatemi 2009). Today, these estimations are done with structural equation modeling, which allows for hypothesis testing, the assessment if these estimates are significantly different from 0 (Neale and Cardon 1992, Medland and Hatemi 2009).

Assumptions of the Classical Twin Design

The classical twin design makes a number of assumptions. It assumes that parental mate choice is random on the phenotype. Violations of this assumption bias heritability estimates downwards. The classic design assumes no gene by gene or gene by environment interactions or correlations. Most of these assumptions are not entirely realistic, though are not more restrictive than assumptions made by any quantitative analysis using methods as simple as multiple regression. For example, anyone who tests a regression without interaction terms also assumes no omitted interactions. Random mating can be assumed with regards to most phenotypes studied, and can also be corrected with an extended design that includes data on parents (Truett et al 1994, Maes et al 1997, Eaves et al 1999, Keller et al 2009, Hatemi et al 2007a). Disregarding this assumption leads to deflated estimates of heritability. Violation of the assumption is not overly problematic, as it only increases the Type II error with regards to genetic heritability, the independent variable that is relatively new and still controversial for the social sciences.

Gene-gene and gene-environment interactions can also be corrected with extensions of the twin design. For example, if a twin modeler establishes there are no common environmental effects on the phenotype, they switch to a model that omits the estimation of the common environment and estimate possible dominance effects (or within-gene interactions, most often abbreviated as D) (Medland and Hatemi 2009). Unfortunately, without phenotypic data from a more extended family structure, only one of these two factors can be included in the model.
Additionally, behavior geneticists often correct for gene-environment interactions. The most commonly used gene-environment models are gene by age interactions (also known as age moderation) models (Feigon et al 2001, Koenig et al 2005, Maes et al 2006, Klump et al 2007). There is a good reason for the widespread use of age as the environmental variable in the interaction. A behavior geneticist would point out that anything social scientists would consider "environmental" could (and probably does) have a heritable component. A gene-environment interaction might not be just a gene-environment interaction but a complex mixture of both gene-gene and gene-environment interactions (Purcell 2002). The only purely environmental

variable without any conceivable mechanism for a heritable component is age. At odds with this, the behavior genetics literature does have rare examples of gene-environment interaction models where the interacted component of the environment is of substantive social scientific interest. Using twin data, Harden et al (2007) modeled the interaction between the heritability of adolescent cognitive aptitude and socioeconomic status.

The issues mentioned in the previous paragraph are not the only reasons gene-environment models are problematic. Apparent but clearly estimatable gene-environment interactions can often be eliminated by mathematical non-linear rescaling of the phenotype's metric. Works as early as 1970 suggest that gene-environment interactions are, most often, measurement artifacts that should be eliminated by metric manipulation (Jinks and Fulker 1970). Item response theory (IRT) models used in this paper to produce a measure of extreme response style are known to eliminate the scaling problem through the operational transformation inherent to the model.

Twin and extended behavioral genetic models do not have to assume no gene-environment interaction as suggested by Charney (2008a). As demonstrated in the cited articles, it is possible to model it. But to do so, a clear theory is required on the mechanism highlighting which element of the environment is interacting with the genes. These environmental variables need to be measured and the interaction between the heritability component and the environment needs to be modeled explicitly. As in the case of all models where independent variables predict a dependent variable, when specific theoretical expectations do not exist for the interaction between the independent variables (in our case the specific aspects of the environment that interacts with the heritable component of the phenotype) it is common to start by assessing the existence of main effects of the independent variables. This is the current study's goal and for this reason it will not venture into the area of gene-environment interactions.

It is not uncommon for a political scientist to *theoretically* consider which independent variables have an impact on the dependent variables. Nobody would criticize a behavioralist claiming that there could be something else influencing the dependent variable as well without offering specifics. A common critique of twin design, on the other hand, argues that the unrealistic assumptions of the model make all findings invalid, but does not specifically highlight specific things that should be corrected for (Charney2008a).

The classical twin design makes the assumption that, on average, the environment MZ co-twins share is equally similar to the environment DZ twins share. This is the equal environment assumption (or EEA). When researchers whose disciplines are founded on the study of environmental influences on behavior (sociology, political science, etc.), criticize the classical twin design, the first target is the equal environment assumption (Horwitz et al 2003a, Richardson and Norgate 2005, Charney 2008a, Beckwith and Morris 2008). The impact of this assumption on the estimates produced by the classical twin design is a legitimate concern. If the equal environment assumption is violated, the classical twin design will overestimate heritability and under-estimate the impact of the common environment. The leverage on estimating heritability stems from having information about how much of their genome, on average, the twins share *and* having information on how much environment, on average, they share. But if we lose the leverage on the environmental similarity in a way that monozygotic twins, on average, share more of their environment than dizygotic twins, we completely lose our ability to accurately estimate heritability.

The EEA was one of Charney's main concerns when criticizing Alford, Funk and Hibbing (Charney2008a). The behavior genetics community, however, has accumulated a large body of relevant literature in the past 30 years that went uncited in the *Perspectives in Politics* exchange. A review of this literature is in order.

Often cited examples of EEA violations include identical twins sharing their bedrooms more often, therefore spending more time with each other than an average fraternal twin pair. Monozygotic twins are also more likely to be dressed alike than dizygotic twins. The use of these arguments was so common in the debate about the methodological appropriateness of the classical twin design that twin researchers started collecting this information (Loehlin and Nichols 1976, Mitchell et al 2007a). But these questions do not constitute the only attempts to measure and compare environmental similarity across MZ and DZ twins. Rose and colleagues measured social contact of co-twins (Rose et al 1988, Rose et al 1990). Kendler and Gardner (1998) used a 12 question battery to measure environmental similarity of twins.

Whether, for example, being dressed alike or not correlates in any way with a given phenotype under study is an important question to consider. In the field of survey research, the question could arise if the characteristic of being dressed alike between ages 0-12 impacts in any way on survey response styles. It would be very difficult to show any mechanism for direct relationship between the two; however, a possible effect of being dressed alike during childhood could cause perception of stronger similarity between the co-twins. While stronger feelings of similarity and closeness between the co-twins could in turn influence many phenotypes under study (such as psychological well being, substance use, etc), a long list of studies failed to find an EEA violation (Hettema et al 1995, Kendler and Gardner 1998, Kendler et al 2000, Xian et al 2000, Derks et al 2006). In light of these findings it is very unlikely that EEA violations would significantly influence survey response styles.

In reality, it is extremely difficult to imagine why parents of twins would socialize one twin differently from the other regardless of how much the twins look alike. A longer laundry list of responses to the EEA critique is available in the Charney debate (Alford et al 2008a, Hannagan and Hatemi 2008, Alford et al 2008b).

But more complex mechanisms should be explored as to how identical twins are treated differently by their parents or wider environment than fraternal twins. Before we consider how zygosity of twins can lead to a different environmental impact on them, we need to explore the accurate identification of the zygosity. While self-identification of zygosity is not problematic (Heath et al 2003, Neale 2003) Scarr and Carter-Saltzman (1979) found that 40% of twins are misclassified by others. A later study put this figure at 19.1% (Kendler et al 1993). Kendler used this information on misclassification to see if it is perceived or actual zygosity that matters. The Kendler model corrected for both actual and perceived zygosity simultaneously (Kendler et al 1993). To date, no evidence has been found to indicate correction for perceived zygosity improves model fit with any phenotype, suggesting the EEA leads to no bias in most cases.

The question of the amount of time twins spend together is more important. Horwitz et al (2003a) pointed out that twins who spend more time together are exposed to similar environmental influences such as peers, friends, their co-twins, etc. But they did not go past identifying this potential problem and considered this as sufficient evidence to dismiss the field of behavioral genetics based on classical twin design. What they should have done was correct for this possible violation within the framework of the behavior genetic models.

What aggravates the problem is that the equal environment assumption is often greatly misunderstood. First of all, the stochastic nature of the assumption is often disregarded. The assumption does not claim that any given identical twin pairs' environment will be the same as any given fraternal twin pairs' environment. Rather, the environmental similarities between the co-twins, *on average*, should be the same for both mono- and dizygotic twin pairs. What is also often forgotten is that any dissimilarity should specifically influence the phenotype studied. The impact of environmental dissimilarity on phenotype should be evaluated directly and theoretically.

While the equal environment assumption of the classical twin design has received a lot of attention in the social sciences, based on the evidence presented in this section it is fair to say that the critique does not invalidate twin research at levels that would make its use pointless. Therefore I proceed with assessing the heritability of survey response styles and call for future theoretical research on the specific possible mechanism that could lead to an assumption violation and empirical research controlling for such identified mechanism. Lacking such specific, theoretically grounded mechanisms this paper will not control for possible effects of equal environment assumption violations. All models tested here will control for age and sex. A separate chapter is devoted to additional control variables identified to have an impact on survey response styles: education and income.

An assumption not discussed in the behavior genetic debates published in the social sciences is the assumption of no measurement error. This assumption stands in most quantitative analysis. For example, it is one of the assumptions of regression models. But while in a regression measurement error only produces type II error, in a behavior genetic model measurement error biases the unique environment estimate upwards producing type I error for this estimate. Heritability and common environmental effects are biased downwards producing type II error. (McCrae and Costa 2003, Riemann et al. 1997). Measurement error will lead to downward bias of co-twin correlations and therefore an inflation of the uncorrelated component of the result, the unique environment. Since all components are expressed in percent of the variance explained an inflation of one component naturally leads to a proportional deflation of the other components. For this reason if measurement of the phenotype is inaccurate that will upward bias the unique environmental component and downward bias the additive genetic and common environment's proportion. A downward bias of a component (towards

0%) also increases the chance that the result will be incorrectly found statistically nonsignificant increasing the chance of Type II error. An upward bias in turn leads to an increased chance of type I error (McCrae and Costa 2003, Riemann et al. 1997).

The best approach to overcoming this problem is extremely careful measurement and the utilization of available techniques to decrease measurement error. As will become apparent from subsequent sections, this study devotes special attention to measurement. Most studies simply take all available items on the surveys for each individual and sum across them but this study will bring in additional considerations of careful measurement that is fairly uncommon for studies assessing survey response styles. In addition to careful measurement, latent variable approaches are used where possible to decrease measurement error further.

The Data

The National Survey of Midlife Development in the United States (MIDUS) is a two-wave general population sample that includes an oversample of twin pairs.[2] Its first wave was collected between 1995-1996. Random digit dialing was used to field a preliminary CATI survey and a follow up self-administered mail survey was sent to all participants. The MIDUS research team utilized aggressive multiple contact and refusal conversion to ensure the representativeness of the sample. A nice pen and

---

[2] In addition to twins, the data also included an oversample of siblings and urban inhabitants. These oversamples were excluded from all analysis in this paper.

$20 of cash was mailed to the phone survey participants. Multiple modes of reminders were presented to the people who received the mail questionnaires. A random sample of nonrespondents were offered $100 to become part of the sample. The sample consisted of respondents between the ages of 25 and 74. For additional information on the sample see Appendix A for an excerpt from the MIDUS website and technical documentation.

The twin oversample of MIDUS was ascertained through approximately 50,000 random digit dialing screening calls where the respondent was asked if they have a twin in the family. This procedure identified 998 twin pairs. The twin oversample contained multiple families with more than one twin pair. After the exclusion twins who were related to other twins in the dataset (to ensure the independence of all co-twins in the data) we were left with 359 monozygotic and 337 same sex dizygotic twins[3]. Different sex twin pairs and twins with uncertain or conflicting zygosity self-report were also excluded from the sample.

A second wave of the panel was fielded 10 years after the initial wave. The second questionnaire contained a large number of questions identical to the first wave. This study only considered questions that were not factual recall questions and appeared in identical form on both questionnaires. The richness of the dataset allowed us to develop the survey response style constructs on the sample of singletons completely independent of the sample of twins used to assess heritability.

---

[3] To maximize sample size one pair of twins were left in the data from families where multiple twin pairs were included in the sample. The decision on which pair to keep in was based on maximizing younger dizygotic same sex male twins in the presented order of importance.

The singleton sample was n=3091 after excluding individuals for whom the specific responses used in this study were not available.

Measurement of Acquiescence

MIDUS included 36 seven point agree-disagree questions (strongly agree, somewhat agree, agree a little, don't know, disagree a little, etc.) that were present on both waves of the MIDUS self-administered questionnaire in completely unmodified form and could not have been skipped due to skip patterns.

Since some of the agree-disagree questions on the questionnaire produced highly uniform results in the entire sample where every individual responded on one end of the scale producing a highly skewed variable. These items could be considered poor, as their responses do not produce much variance. Also, when one end of the scale (for example the agreement end) is just absolutely unsupportable people who would normally acquiesce might not do so. This is true vice versa for questions skewed in the direction of agreement. Highly skewed items introduce measurement error. The inclusion of these items in the acquiescence construct is problematic; they bias the mean independent of any respondent characteristic, only depending on the direction of the skew. To minimize such bias, questions that produced highly skewed responses in wave I of the singleton sample were removed. If the absolute value of skewness was above 1 the question was excluded from the analysis leaving us with 19 eligible items. (See Appendix B for the 19 specific items.)

Literature operationalizing acquiescence using secondary analysis of surveys either used the sum of "clear" agreements minus the sum of "clear" disagreements or took the difference between all agreeing and all disagreeing responses. Van Herk et al (2004) used 1s and 2s minus 8s and 9s of a 9 point agree-disagree scale, but for a 5 point scale, much like Bachman and O'Malley (1984), used 1 and 2 minus 4 and 5. The most appropriate operationalization is debatable but lacking a clear theory on the best measure of acquiescence it is difficult to reconcile this debate.

Past studies also assessed agreement and disagreement as distinct constructs (Bachman and O'Malley 1984, Baumgartner and Steenkamp 2001). Bachman and O'Malley (1984) found that agreement and disagreement counterintuitively correlated positively, suggesting that acquiescence and extreme response are constructs that independently better capture survey response styles. Directional bias approach that averages the raw agree-disagree scale responses is also prevalent in the literature (Hui and Triandis 1985, Gilman et al 2008, Spector 1987). This measure is popular amongst scholars in search of an acquiescence correction (Hofstede 1980, Au and Cheung 2002, Ayçan et al 2000, Hofstede, 2001, Leung and Bond 1989, Morris et al 1998, Schwartz 1994, Smith et al 2002) but such easy fixes are not uncontroversial (Smith 2004). This measurement could also be heavily biased by original question content if true scores of the items are, on average, correlated.[4]

[4] In an attempt to control for this problem I considered using uncorrelated items. But unfortunately it is impossible to separate correlations arising from the item true scores and the acquiescence bias. This consideration was abandoned.

In this study I will assess the most comprehensive measure of acquiescence, or *net acquiescence response style* as described by Baungartner and Steenkamp (2001) also used by Bachman and O'Malley (1984) and van Herk et al (2004) that incorporate both agreement and disagreement in a single construct by taking their difference. To preempt additional debates on operationalization I compiled three different measures of acquiescence. All 19 questions used in producing the construct were agree-disagree questions measured on a seven point scale. Measure 1 (acq1) sums the number of strongly agree responses across the 19 questions for each individual and subtracts the sum of strongly disagree responses. Measure 2 (acq2) sums the strongly and somewhat agrees and subtracts the corresponding disagrees. And finally, acq3 sums across all agreeing responses and subtracts all disagreeing responses.

The operationalization as laid out here uses a simple sum across items. A better approach would be to use a latent variable model where measurement error is removed from the construct. (Such an approach is presented for extreme response in the next section.) Unfortunately the net acquiescence response style approach makes use of the sum of two positively correlated constructs (agreement and disagreement) that prevents the utilization of a single latent variable. The one approach in the literature that was able to overcome this problem using a measurement error free latent variable model of acquiescence requires the utilization of a balanced set of questions with half of the indicators using reversed items (Billiet and McClendon 2000, Cheung and Rensvold 2000, van de Vijver and

Leung, 1997)[5]. This approach was not possible with the available data. Due to the unavailability of latent variable approaches traditionally used to minimize measurement error I suspect that the reliability estimates will be lower then they would be had we have a better, more measurement error free, way to construct an acquiescence score. Due to this inability to more carefully measure acquiescence, additive genetic and common environmental component estimates could be artificially deflated and the unique environmental component estimate could be artificially inflated. Again, this is due to the higher levels of random measurement error deflating the co-twin correlations and increasing the uncorrelated component of the model.

Measurement of Extreme Response Style

One hundred and thirty-four ordinal response category questions were identified as present on both waves of the MIDUS self-administered questionnaire in completely unmodified form. This list excluded factual recall questions, questions that could have been skipped due to skip patterns and personality items that are used as frame of reference for the results. To be included the ordinal questions needed to have at least four ordinal response categories and no additional category offered beyond the ordinal response options.

While most studies used all available questions on the survey to produce a measure of extreme response, Greenleaf (1992) convincingly argued that such an

[5] Experimental studies also often utilized item reversal (Landsberger and Saavedra 1967, Javeline 1999)

approach could be improved on to minimize measurement error. Greenleaf (1992) argues that the best items to measure extreme response (1) have low item – extreme response correlations (in other words they do not have large frequencies at one of the extremes), (2) are uncorrelated with each other and (3) have roughly equal extreme response proportions.

To identify the ideal questions for measuring extreme response style, first, the 134 questions were correlated with their own dichotomous extreme response score using Spearman's correlation on the wave I singleton data. A response was considered extreme if it was at either one of the ends of the ordinal scale. Questions where the extreme response correlated with the raw response above an absolute value of 0.3 were eliminated, leaving 21 questions. The questions eliminated were predominantly questions with highly skewed responses.

The second step in the selection process was to ensure that the responses to these questions did not correlate with each other. The 21 questions were entered into an exploratory factor analysis with orthogonal (varimax) rotation. This identified strongly correlated items (usually items from the same battery). For correlated item-pairs the ones with higher extreme response - response correlations were excluded removing seven items. Spearman's correlation matrix of the remaining 14 items was evaluated and two additional items were eliminated due to correlations above 0.3.

For the remaining 12 items extreme response proportion ranged from 5.8% to 31.1%. Greenleaf (1992) calls for roughly equal extreme response proportions across the items. The four items with the lowest extreme response rate were eliminated leaving 8 items with the narrower range of 14.7% to 31.1%. Greenleaf's (1992) article ended up with lower correlations and a better extreme response range for his dataset. Unfortunately comparable numbers could not have been achieved with MIDUS. But to improve on Greenleaf's approach a latent variable IRT model is used to eliminate additional measurement error from the extreme response style construct. In an IRT model multiple categorical indicators are used to produce a latent factor underlying these indicators. This latent factor derives common variation from the indicators discarding variation unique to the individual items and therefore produces a more measurement error-free construct than what a simple summation of the indicators would produce (Ostini and Nering 2005).

En sum, Greenleaf's (1992) rules were used on the wave I singleton sample to identify the 8 best indicators of extreme response (also listed in Appendix C). These questions had Inter-Item and Item-ERS Spearman's correlations under 0.3, and extreme response proportions between 14.7% and 31.1%. Results were comparable in wave II of the singleton sample and in both waves of the twin sample. As described above responses were recoded into dichotomous extreme response items coding the extremes of the ordinal scales as 1 and all other responses as 0. Contrary to acquiescence where a latent variable approach was not strait forward, these extreme response items can be used as indicators of a latent variable measuring extreme response style. An IRT model was used to construct a continuous measure of extreme response style (de Jong et al 2008).

Given the larger pool of available items to select the best measures and the appropriate latent variable approach for measuring extreme response style, I expect that the measure of extreme response will be more reliable then the measures of acquiescence and the ACE variance decomposition estimates will be more accurate.

Methodology

Data cleaning, descriptive statistics, correlations and the exploratory factor analysis for item selection were calculated using SPSS. All other models presented were estimated with Mplus (Muthen and Muthen 2007) using maximum likelihood unless otherwise stated. Since not all latent variables are created equal, the uncertainty of the estimated latents was accounted for by the simultaneous estimation of the measurement (IRT) model with every other presented result. For this reason computationally intensive numerical integration was required. While it is customary to present bootstrapped or likelihood ratio based confidence intervals for the ACE components, this is computationally too taxing for latent variable models such as the one presented here.[6] In these situations it is common to present standard errors based significance levels (Boomsma et al 2007, Muthen et al 2006).

Reliability of Extreme Response and Acquiescence

Due to the need to gain leverage on measurement error it is important to test the reliabilities of both acquiescence and extreme response styles. Several studies

[6] It is not uncommon for these models to run for 5 to 10 hours on a Core 2 Quad, 4 CPU PC running with 8GB of RAM. Hundreds of bootstrapped replications of these models are simply not feasible computationally at this time.

pointed to the test-retest stability of both survey response styles in panel design (Berg and Collier 1953, Bachman and O'Malley 1984). They attributed the different survey response styles to different personalities. Hui and Triandis (1985) took issue with this showing that stability decreases (though still high) as time passes between measurements. Based on this and the careful measurement of both constructs we can expect that acquiescence and survey response styles will have reasonable internal and test-retest reliability even with 10 years between measurements. Since a latent variable approach that decreases measurement error is only used for extreme response it is expected to have better reliability.

Using the first wave of MIDUS, internal reliability of the extreme response measure was assessed through the IRT construction of two latents, one using the first four and the other using the second four extreme response indicator items (see Figure 1). The IRT model and the correlation of latents were estimated simultaneously. Correlation of the latents was r=0.608 (SE=0.029). To compare, the well established and understood big 5 personality construct of openness, the only personality that was measured with at least 6 items on MIDUS (the minimum for identification of the internal reliability IRT model), has an internal reliability of r=0.574 (SE=0.010)[7].

Using the second wave of MIDUS, test-retest reliability was also assessed (see Figure 2). The test model assumed that the indicators contribute equally to the latent extreme response variable in both waves, therefore we can restrict the loadings to equal each other in the two waves. These assumptions were tested

[7] Internal reliabilities of openness was estimated with weighted least squares (WLS) due to convergence issues with the equivalent maximum likelihood (ML) models.

empirically; the restricted model fit better then the unrestricted model.[8] Extreme response test-retest reliability was r=0.833 (SE=0.029). For the big 5 personality items this was: Agreeableness r=0.795 (SE=0.015), Conscientiousness r=0.958 (SE=0.02), Extraversion r=0.854, (SE=0.013) Neuroticism r=0.726 (SE=0.018), Openness r=0.855 (SE=0.013).[9] While this analysis produced higher test-retest reliability for some personality measures, personality batteries had the unfair advantage of having multiple category response options. Even independent of this handicap, extreme response internal and test-retest reliability can be considered comparable to the well-established Big 5 personality items.

The reliability test of acquiescence was more strait forward as it does not use a latent variable model. As with extreme response, internal reliability was tested using wave I of MIDUS and test-retest reliability was tested using both MIDUS waves. As mentioned above three different measures of acquiescence are used, dubbed acq1, acq2 and acq3. Again, acq1 subtracted the number of strongly disagreeing responses from the strongly agreeing ones to calculate acquiescence; acq2 used strongly and somewhat agreeing and disagreeing responses while acq3 used all agreeing and all disagreeing responses.

To test internal reliability half of the 19 items (odd numbered) were used to calculate a score of acquiescence and these were correlated to the other half (even

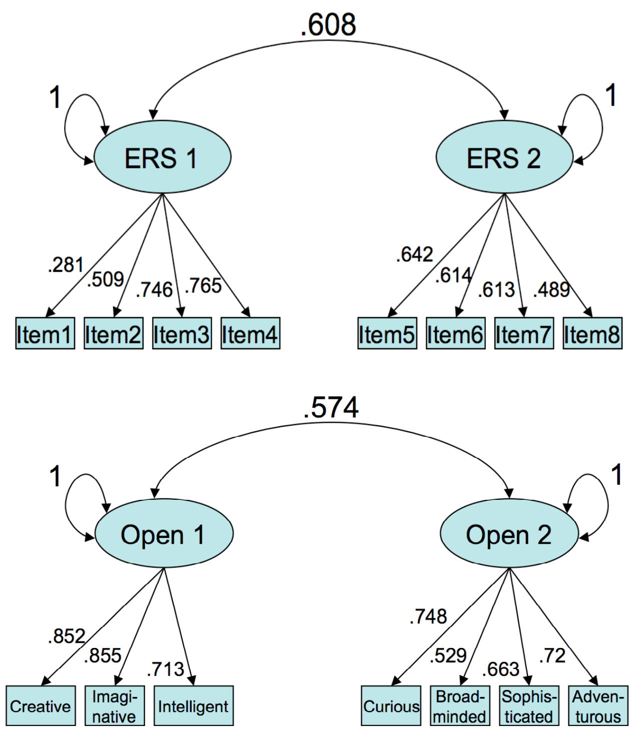
---

8 Unrestricted model: AIC=35314.262, BIC=35513.447
Restricted model: AIC=35304.919, BIC=35407.530

9 All loadings for these models were significant at $p<0.001$.

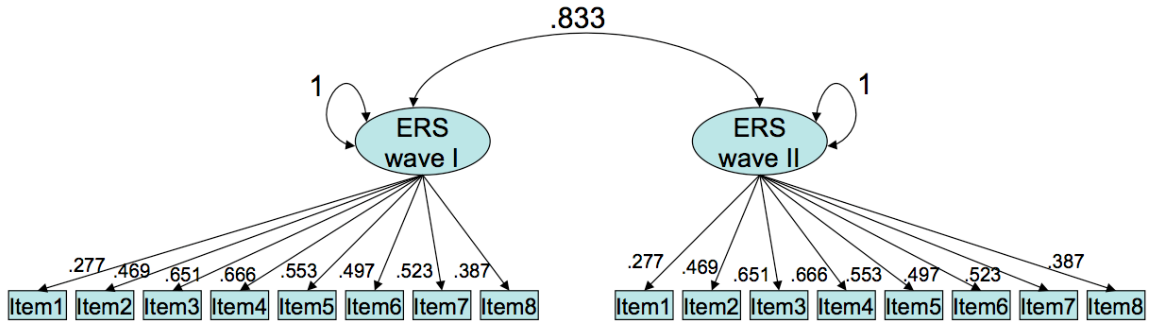
numbered) items.[10] Internal reliability of acq1 was 0.573, acq2 was 0.55 and acq3 was 0.537 all significant at ($p<0.001$). As expected, these reliabilities are lower than extreme response reliabilities and are comparable to the internal reliability of openness. Test-retest reliabilities were also in the 0.5 range (acq1=0.544, acq2=0.527, acq3=0.512, all $p<0.001$).

Figure 1: Measurement models testing the internal reliability of extreme response and openness (Big 5 personality construct with the largest number of indicators)



[10] Item 19 was dropped to ensure that the two halves were using the same number of items. Sensitivity analysis dropping one different item revealed very similar results. Internal reliability correlations for acquiescence are Pearson correlations.

Figure 2: Test Retest Reliability of Extreme Response Style[11]



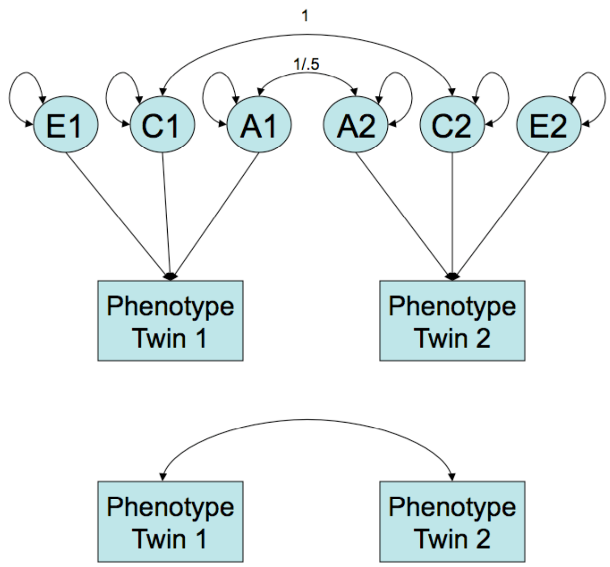
ACE Decomposition of Variance

Before I introduce the specific model used in this study, it is important to give an overview of the simple univariate ACE model. Figure 3 graphically shows the classical twin structural equation model for a single phenotype and the baseline model it is traditionally compared to. This comparison is done with a chi-square difference test and assesses the appropriateness of using the ACE model. Both models are two-group models. The more complicated ACE model introduces different model restrictions for MZ and DZ twins. The three circles represent latent factors of additive genetic influence (A), common environment (C) and unique environment (E). These latent factors influence the score of the phenotype as denoted by the path going from the latent circles to the observed phenotype. The

[11] I realize item 1 does not appear to measure extreme response well since it does not reach the loading of the minimum of .3 used in exploratory factor models, but rather than engage in a post-hoc data driven modification of the model, I decided to keep the item. One possible reason this item does not measure well is that it was very early in the long survey when respondents are still relatively attentive.

squared standardized path estimates for these arrows give us the proportion of the latent factors' impact.

Figure 3: Univariate ACE model and its baseline model below it.



A, C and E are not measured directly but assumptions are made about their relations to one another. We know that monozygotic twins share 100% of their genome and therefore the latent factor of A for twin 1 and twin 2 are perfectly correlated for monozygotic twins as denoted by the curved arrow and the number 1 above it. We know that dizygotic twins, on average, share 50% of their genome and for this reason the latent factor of A for twin 1 and twin 2 are correlated 0.5 for dizygotic twins. Because we assume EEA as discussed above in detail, the correlation between C1 and C2 is fixed at 1 for both MZ and DZ twins. E1 and E2 are uncorrelated.

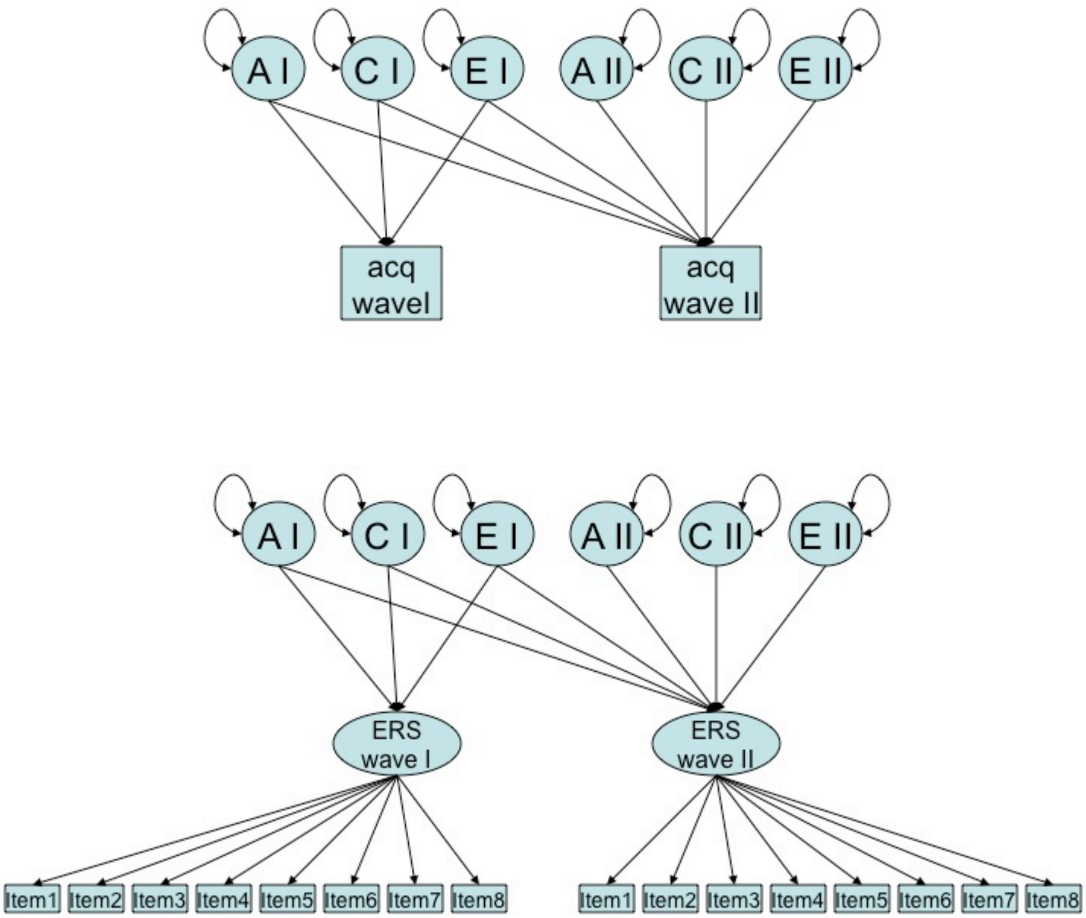
In most behavior genetics studies, follow-up models restrict the nonsignificant paths from A, C and E to the phenotype to 0. But Medland and

Hatemi (2009) warn against the use of such follow up models when the full model is estimated on data with small, underpowered samples. When sample sizes are small, results are often nonsignificant due to lack of power (type II error) and not because the effect is not present in the population. Inappropriately restricting a path to 0 in a model will bias the results and for this reason it is best to keep the effects of heritability, common environment and unique environment in the model, even if one of these is nonsignificant.

The phenotype in these models does not necessarily have to be the raw measured scores of construct of interest. They can be corrected for independent variables leaving only the post-control residual variance to be decomposed into A, C and E. Generally age and sex are controlled for in these models. This thesis will also include a section discussing the impact of income and education as predictors. Baseline models are also adjusted accordingly with the predictors.

Since panel data is available it is more advantageous to use a multivariate model that uses both waves of the panel. This is important as the sample size at hand is considered small and is quite underpowered to estimate heritability and common environment (but not unique environment) by twin research standards. (Post-estimation power analyses are presented with the rest of the results.) Behavior genetic studies often use sample sizes in the 10,000's. Adding the additional data from wave II can increase power and can enlighten us concerning the source of test-retest stability of extreme response styles.

Figure 4: Cholesky Decomposition ACE model for Acquiescence and Extreme Response (simultaneously incorporating the IRT and ACE component of the model)



The most commonly used multivariate ACE model is the Cholesky decomposition ACE model. In its bivariate form (the form used in this study where two phenotypes are assessed) the model estimates the additive genetic, common and unique environmental effects of the first phenotype, decomposes the covariation between the two phenotypes into covariation due to additive genetic, common and unique environmental covariation and estimates the additive genetic, common and unique environmental effects of the second phenotype's leftover variation not explained by the covariation with first phenotype. In the baseline model for the Cholesky decomposition the four variables (wave1 for twin1, wave2 for twin1, wave1 for twin2 and wave2 for twin2) are corrected for the predictors and are all correlated with each other. Due to size limitations Figure 4 only shows the heritability model and for just twin 1. Same restrictions on the covariation between twin1 and 2 presented in the univariate model are in place for A, C and E.

Since the two phenotypes in this study are the same, just measured 10 years apart, it is reasonable to add an additional restriction to the Cholesky decomposition ACE model. It is reasonable to assume that the impact of heritability, common environment and unique environment will be identical across the two waves. The model fit statistics presented in Appendix D show that the restricted models that include the ACE restrictions and the additional restrictions equating the impact of heritability, common and the unique environment across the two waves do not hamper model fit as compared to the unrestricted baseline models.

Results

Table 1 present the heritability of variance and test-retest covariance for the three measures of acquiescence and for extreme response style. The results from the three measures of acquiescence corroborate well showing significant 25% to 34% heritability, small and nonsignificant (10%-11%) common environmental effect and a large and highly significant (55%-64%) unique environmental effect. Note that any measurement error present biases this latter result upwards. Had we measured acquiescence without measurement error we could expect this effect to be lower and the other two effects to increase proportionally (McCrae and Costa 2003, Riemann et al. 1997).

Table 1: ACE decomposition of phenotype variance and covariance. Standard Errors in Parentheses

	Acq1		Acq2		Acq3		ERS	
Variance								
Heritability (A)	0.251	*	0.341	**	0.294	**	0.195	
	(0.117)		(0.108)		(0.113)		(0.195)	
Common Environment (C)	0.109		0.113		0.102		0.324	*
	(0.1)		(0.094)		(0.096)		(0.161)	
Unique Environment (E)	0.64	***	0.546	***	0.605	***	0.481	***
	(0.038)		(0.034)		(0.037)		(0.069)	
Covariance Between Wave I and II								
Heritability (A)	0.251	*	0.341	**	0.294	**	0.195	
	(0.117)		(0.108)		(0.113)		(0.195)	
Common Environment (C)	0.109		0.113		0.102		0.314	+
	(0.1)		(0.094)		(0.096)		(0.161)	
Unique Environment (E)	0.222	***	0.149	***	0.174	***	0.291	***
	(0.039)		(0.033)		(0.037)		(0.075)	
Total Covariance	0.582	***	0.603	***	0.569	***	0.8	***
	(0.023)		(0.022)		(0.024)		(0.037)	
Controls Variables								
Female	0.03		0.09	**	0.101	***	0.172	***
	(0.03)		(0.03)		(0.029)		(0.037)	
Age	0.013		0.154	***	0.164	***	0.235	***
	(0.02)		(0.03)		(0.03)		(0.033)	

+ p<0.10, * p<0.05, ** p<0.01, *** p<0.001

Notes: Results are all standardized. Extreme response loadings are sightly different once age and sex are controlled for. Rough magnitude and significance of the loadings have not changed and therefore are not reported. Discrepancies observed between the test-retest reliability correlation and the total covariance presented here are due to the inclusion of control variables.

It is interesting that extreme response, a survey response trait that has been assessed to result from the same processes with other survey response styles like acquiescence, provide different results. The trait is not significantly heritable (20%) even with the improved leverage on measurement error. Common environmental effects, on the other hand, are significant at 32%. It is possible that this nonsignificant result for heritability is due to lack of power as twin designs tends to be extremely underpowered in most instances. Assuming the point estimates for extreme response are accurate for the population (A=0.2, C=0.32 and E=0.48) with the given panel design, the available sample size and assuming no item missing data, no attrition and perfectly normally distributed phenotype I only had 59.2% power to detect a significant relationship[12]. In reality power is probably somewhat lower due to the unrealistic nature of these assumptions.

Additionally, genetic and common environmental effects heavily drive the covariation between the two traits. Using the example of the first measure of acquiescence, the numbers presented in the table can be interpreted as: of the 0.582 correlation, 0.251 is driven by heritability (A), 0.109 is driven by the common environment (C - though this result is not significant) and 0.222 is driven by the unique environment (E). Expressed in percentages 0.251/0.582=43% of the covariation is driven by A, 19% is driven by C and 38% is driven by E. For acq2 these percents of covariation are A=57%, C=19% and E=25%; for acq3 A=52%, C=18% and E=31%; and for extreme response style A=24%, C=39% and E=36%.

Note that with the exception of extreme response where there is a slight (1%) deviation for common environmental covariation (which also drove down

---

[12] Power to detect C was above 80%. Power analysis for the acquiescence measured yielded more power to detect A significantly but still under 80% for all three models. For acquiescence power to detect C was extremely low. Behavior geneticists generally do not consider 10% variance explained by a single component other then E as worth exploring because of the lack of power, the enormous sample size required to achieve significant results. Power to detect E was always within rounding distance of a 100%.

significance to $p<0.10$ from $p<0.05$), the amount of variance and covariance driven by the heritability and common environmental components is equal. Genes and common environmental components responsible for variation in the survey response trait are also equally responsible for the stability of these traits in test-retest situations.

As for the control variables, with the exception of the first measure of acquiescence where results were nonsignificant, females are significantly more likely to acquiesce and provide an extreme response. The same is true for older respondents.

Additional Control Variables

As described above with regards to the debates of the trade, behavioral geneticists are wary of controlling for additional variables social scientists would naturally consider. This is because these "environmental" variables are often not fully environmental and have a heritable component (Purcell 2002).

If we control for, extract variance from, the phenotype attributed to such "environmental" controls we will only bias (and make meaningless) the heritability, common and unique environmental estimates of the phenotype of interest. But in the social sciences it is customary to control for known covariates as we are interested in the sources of the unexplained variance after control variables are included. Technically it is possible to extract variance attributed to controls and decompose the leftover variance into additive genetic, common and unique environmental sources. This is what the present section will do. Education has often been used as a predictor of survey response styles. It has been considered a proxy for social status (Lenski and Leggett 1960, Landsberger and Saavedra 1967, Carr and Kraus 1978). In addition to education, income could also serve as a proxy for social status. Table 2 presents the ACE variance decomposition results with income and education corrected for (along with age and sex already controlled for in the model presented in table 1). Due to the high colinearity of these traits they will be considered independently in the models.

As seen from Table 2 additional controls do not have much substantive or statistical effect on the heritability, common and unique environmental impact on acquiescence or extreme response. Income or education correction also does not really change the impact of age or sex on the phenotype. As found in previous studies income and education has negative impact on acquiescence and extreme response, though income's effect on extreme response is not significant.

Table 2: ACE decomposition of phenotype variance and covariance with additional controls. Standard errors in parentheses.

	Acq1 ct ed		Acq1 ct inc		Acq2 ct ed		Acq2 ct inc		Acq3 ct ed		Acq3 ct inc		ERS ct ed		*ERS ct inc*	
Variance																
Heritability (A)	0.258	*	0.239	*	0.288	*	0.34	**	0.211	+	0.28	*	0.187		0.2	
	(0.119)		(0.119)		(0.114)		(0.11)		(0.118)		(0.115)		(0.199)		(0.196)	
Common Environment (C)	0.085		0.104		0.101		0.096		0.114		0.088		0.308	+	0.309	+
	(0.101)		(0.1)		(0.098)		(0.095)		(0.099)		(0.097)		(0.164)		(0.161)	
Unique Environment (E)	0.657	***	0.657	***	0.612	***	0.564	***	0.675	***	0.632	***	0.505	***	0.491	***
	(0.038)		(0.039)		(0.037)		(0.035)		(0.039)		(0.038)		(0.07)		(0.07)	
Covariance Between Wave I and II																
Heritability (A)	0.258	*	0.239	*	0.288	*	0.34	**	0.211	+	0.28	**	0.187		0.2	
	(0.119)		(0.119)		(0.114)		(0.11)		(0.118)		(0.115)		(0.199)		(0.196)	
Common Environment (C)	0.085		0.104		0.101		0.096		0.114		0.088		0.306	+	0.302	+
	(0.101)		(0.1)		(0.098)		(0.095)		(0.099)		(0.097)		(0.164)		(0.161)	
Unique Environment (E)	0.226	***	0.228	***	0.175	***	0.147	***	0.202	***	0.178	***	0.289	***	0.283	***
	(0.039)		(0.04)		(0.037)		(0.034)		(0.04)		(0.039)		(0.077)		(0.077)	
Total Covariance	0.568	***	0.571	***	0.564	***	0.583	***	0.528	***	0.546	***	0.782	***	0.784	***
	(0.023)		(0.023)		(0.024)		(0.023)		(0.025)		(0.025)		(0.043)		(0.043)	
Controls Variables																
Female	0.008		-0.038		0.059	*	0.004		0.072	**	0.013		0.156	***	0.158	***
	(0.029)		(0.032)		(0.029)		(0.032)		(0.028)		(0.031)		(0.035)		(0.04)	
Age	-0.01		-0.001		0.121	***	0.136	***	0.124	***	0.14	***	0.217	***	0.239	***
	(0.029)		(0.029)		(0.028)		(0.029)		(0.027)		(0.028)		(0.037)		(0.035)	
Education	-0.173	***			-0.26	***			-0.265	***			-0.142	***		
	(0.026)				(0.026)				(0.025)				(0.034)			
Income			-0.148	***			-0.192	***			-0.202	***			-0.039	
			(0.026)				(0.026)				(0.026)				(0.034)	

+ p<0.10, * p<0.05, ** p<0.01, *** p<0.001

Notes: Results are all standardized.

Discussion

The first striking aspect of the results presented is the difference between acquiescence and extreme response style. While no significant additive genetic effect has been found in extreme response, acquiescence appears to be heritable. This does not mean that extreme response have no heritable component, it just means we could not find a statistically significant additive genetic component with this admittedly underpowered sample. Especially given the suspected amount of measurement error in the two traits, results are suggestive as to the magnitude of heritability in extreme response and acquiescence. For common environment the story is vice versa where extreme response appears to have a significant common environmental influence while the C estimate for acquiescence is too low to even consider by behavior genetics standards.

The survey response literature considered these traits together as survey response styles (or survey response sets). Implicit differentiation only emerged from the acquiescence literature that did not list extreme response as a form of satisficing behavior (Krosnick 1991). According to these findings the underlying mechanisms driving these traits are probably different and call for a more nuanced discussion of these traits.

The Cholesky design using the panel data allowed for the analysis of sources of trait stability. It appears that additive genetic and common environmental variation heavily contributes to trait stability. At first glance this is intuitive. People's genes do not change over time and common environmental, based on its most simple interpretation is rooted in socialization, something both co-twin received equally.

It appears that the satisficing literature is more in line with the findings. First of all, it considers acquiescence as a source of satisficing but extreme response is not. This could explain the apparent differences in the identified sources of differential variation. It can be stated with confidence that any socialization factors are negligible in interpersonal differences of acquiescence.

Looking at the heritability components of acquiescence there are multiple mechanisms though which the relationship between a person's DNA and acquiescence could emerge. Krosnick identified four such mechanisms (Krosnick 1991, Korsnick 1999). Primarily, taking acquiescence at its face value, the most simple mechanism is that it could be driven by the big 5 personality trait, (1) agreeableness. In addition people who acquiesce are probably (2) more extraverted and (3) more open. All three of these five big 5 personality traits are genetically heritable (Jang et al 1998). Secondly, and still staying with big 5 personality correlates, when considering that acquiescence could emerge from satisficing behavior, people who understand that they are accountable for their survey responses are less likely to satisfice and therefore less likely to acquiesce (Krosnick 1991). Conscientious individuals are more likely to be more accountable. Also, neurotic individuals are more likely to fear being held accountable so they act accordingly and minimize acquiescence. Third, less satisficing has been tied to higher cognitive ability (Krosnick 1991), a highly heritable trait (Plomin et al 1994, Rietveld et al 2003). And finally less satisficing has also been tied to higher levels of altruism (Krosnick 1991), once again a heritable trait (Rushton et al 1986).

Survey response styles (sets) have been treated in the past as personality traits (Lentz 1938). It is plausible that the analysis of a more powerful sample could yield significant heritable component for extreme response. If so, comprehensive measures of personality dimensions, such as the Big 5, would be a good place to start in finding specific genetic bases for extreme response, but the expectations for these associations are also less clear.

Future Research

The classical twin design, though important, can only provide us with very little information on what specific variables are driving individual differences. They can point us in various directions. The prescriptions for acquiescence and extreme response are different due to the apparently different results. For acquiescence one place to look for specific predictors is in the genome. Among the studies of a large number of phenotypes, association studies have successfully identified several genes with an impact on social, and political behavior (Fowler and Dawes 2008, Dawes and Fowler 2009). The same association techniques could be used to find genes with an impact on acquiescence. While a genome wide approach would be appropriate, these studies suffer from lack of power and an inductive, exploratory approach that risks type I error. A better initial deductive test would be finding genes already associated with the theoretically plausible heritable correlates:

personality, cognitive ability and altruism. For example, the database, www.genecards.org (accessed on November 1, 2009) produced 464 articles citing 137 specific genetic markers when searching for "personality". Of these markers 45 were cited at least twice, 30 at least three, 21 at least four and 15 at least five times. These better-corroborated genetic markers of various measures of personality could serve as testable hypotheses in an association study.

For extreme response the search should focus on environmental predictors that co-twins could share over their lifetime. An easy target for such research is childhood familial and social variables and styles of socialization. To date no study explored such variables as predictors of extreme response.

Needless to say, these results should be corroborated, preferably with larger, more powerful datasets, before anything can be stated with reasonable certainty. A more powerful sample could exhibit a significant heritable component of extreme response, for example. But also, results might be different when different modes are used or when different cultures are evaluated. The current design could be extended to incorporate other family members for more precise estimation. Gene-environment interactions could be tested and there is room for both theoretical and empirical work on possible equal environment assumption violations inappropriately influencing the results presented here.

Other aspects of survey response could be explored as well. In addition to acquiescence and extreme response, I considered estimating the impact of nonresponse and respondent uncertainty.

Initially, the data appeared to be an excellent candidate for the assessment nonresponse. While no individual level meta-data was available on initial sampling, number of contacts, success of contact, it did have information on people who were surveyed by the phone and lost later in the self-administered questionnaire. Unfortunately for the sample of twins, this number was extremely low. This makes sense, as twins usually realize that they are important to science and tend to respond at higher rates.[13] Inconsistencies between the response rate of different and same sex dizygotic twins lead me to believe that more aggressive refusal conversion strategy was used for same sex twins, the twins most often used by twin designs. Unfortunately, sources of this discrepancy were not documented by the MIDUS project. In addition, the documentation contained no information on how the twin oversampling was done making the use of response rate as a phenotype practically impossible.

A second attempt was made to use nonresponse in wave II of the survey. Unfortunately this nonresponse could occur for various reasons such as death, researchers lost track, etc. While some documentation is available for reasons of non-contact, it is largely incomplete as it only documents failed recontact attempts. But in some cases recontact was not even attempted since the respondent provided no recontact information.

[13] Actually, monozygotic twins are more likely to believe they are important to science when in reality dizygotic twins are more important for twin samples. An ideal sample to estimate heritability contains more dizygotic twins. Different response rates stemming from this differential could be a source of EEA violation treating nonresponse as the phenotype. This possibility deserves more attention.

There is one possibility of gaining leverage on nonresponse. The first wave of the survey included one page of extensive recontact details that was to be used for the second wave of the study. The MIDUS dataset does contain information on the status of this sheet for people who returned the self-administered mail questionnaire. This information can be used as a proxy for participation willingness. While variance is low, as less then 10% of the twins failed to fill out or return this recontact sheet, preliminary analysis using a logistic univariate ACE model (much like the one presented in Figure 3) suggests that participation willingness has absolutely no heritable component and is significantly driven by common environment (47%, $p<0.05$) and unique environment (53%, $p<0.001$). Although this information is not conclusive due to the small sample, indirect measure of willingness to participate and the heavily selected sample already excluding people who did not pick up the phone, refusing to participate in the phone survey or blowing off the self-administered follow up study completely. If these results are corroborated on other samples, possibly with better measures, it contributes to evidence for the lack of heritability of survey participation.

Building on current findings multivariate genetic assessments are in order to test common genetic covariation between acquiescence and its predictors: (1) Big 5 personality measures, (2) cognitive ability and (3) altruism. If genetic covariation is indentified direct association should be tested between the genes already linked to these three predictors and acquiescence.

Works Cited

Alford, J. R., Funk, C. L., and Hibbing, J. R. (2005). Are political orientations genetically transmitted? *American Political Science Review*, 99(2):153–167.

Alford, J. R., Funk, C. L., and Hibbing, J. R. (2008a). Beyond liberals and conservatives to political genotypes and phenotypes. *Perspectives on Politics*, 6(2):321–328.

Alford, J. R., Funk, C. L., and Hibbing, J. R. (2008b). Twin studies, molecular genetics, politics, and tolerance: A response to Beckwith and Morris. *Perspectives on Politics,* 6(4):793–797.

Arce-Ferrer, A. J. (2006). An investigation into the factors influencing extreme-response style: Improving meaning of translated and culturally adapted rating scales. *Educational and Psychological Measurement*, 66(3):374–392.

Au, K. and Cheung, M. W. L. (2002). Consequences of intra cultural variation: An exploratory study of job autonomy in 42 nations. In Boski, P., van de Vijver, F. J. R., and Chodynicka, A. M., editors, New directions in cross-cultural psychology. Polish Psychological Association.

Aycan, Z., Kanungo, R. N., Mendonca, M., Yu, K., Deller, J., Stahl, G., and Kurshid, A. (2000). Impact of culture on human resource management practices: A 10-country comparison*. Applied Psychology An International Review*, 49(1):192–221.

Ayidiya, S. A. and McClendon, M. J. (1990). Response effects in mail surveys. *Public Opinion Quarterly,* 54(2):229–247.

Bachman, J. G. and O'Malley, P. M. (1984). Yea-saying, nay-saying, and going to extremes: Black-white differences in response styles. *Public Opinion Quarterly*, 48(2):491–509.

Baumgartner, H. and Steenkamp, J.-B. E. M. (2001). Response styles in marketing research: A cross-national investigation*. Journal of Market Research,* 38:143–156.

Beckwith, J. and Morris, C. A. (2008). Twin studies of political behavior: Untenable assumptions? *Perspectives on Politics*, 6(4):785–791.

Belli, R. F., Herzog, A. R., and VanHoewyk, J. (1999). Scale simplification of expectations for survival: Cognitive ability and the quality of survey responses. *Cognitive Technology*, 4(2):29–38.

Berg, I. A. and Collier, J. S. (1953). Personality and group differences in extreme response sets. *Educational and Psychological Measurement*, 13:164–169.

Biemer, P. P. (2001). Nonresponse bias and measurement bias in a comparison of face to face and telephone interviewing. *Journal of Official Statistics*, 17(2):295–320.

Billiet, J. (2003). Cross-cultural equivalence with structural equation modeling. In Harkness, J. A., Van de Vijver, F. J. R., and Mohler, P. P., editors, Cross-Cultural Survey Methods, Wiley Series in Survey Methodology, chapter 16, pages 247–263. Wiley and Sons.

Billiet, J. and McClendon, M. J. (2000). Modeling acquiescence in measurement models for two balanced sets of items. *Structural Equation Modeling: A Multidisciplinary Journal*, 7(4):608–628.

Boomsma, D., Cacioppo, J., Muth´en, B., Asparouhov, T., , and Clark, S. (2007). Longitudinal genetic analysis for loneliness in dutch twins. *Twin Research and Human Genetics*, 10:267–273.

Bouchard, T. J., Lykken, D. T., McGue, M., Segal, N. L., and Tellegen, A. (1990). Sources of human psychological differences: The minnesota study of twins reared apart. *Science*, 250:223– 228.

Burkhart, R. E. and Lewis-Beck, M. S. (1994). Comparative democracy: The economic development thesis. *The American Political Science Review,* 88(4):903–910.

Carr, L. G. and Krause, N. (1978). Social status, psychiatric symptomatology, and response bias. *Journal of Health and Social Behavior,* 19(1):86–91.

Cesarini, D., Dawes, C. T., Fowler, J., Johannesson, M., Lichtenstein, P., and Wallace, B. (2008). Heritability of cooperative behavior in the trust game*. Proceedings of the National Academy of Sciences*, 105(10):3721–3726.

Cesarini, D., Dawes, C. T., Johannesson, M., Lichtenstein, P., and Wallace, B. (2009). Genetic variation in preferences for giving and risk taking. *Quarterly Journal of Economics*, 124(2):809–842.

Charney, E. (2008a). Genes and ideologies. *Perspectives on Politics*, 6(2):299–319.

Charney, E. (2008b). Politics, genetics, and "greedy reductionism". *Perspectives on Politics*, 6(2):337–343.

Chen, C., ying Lee, S., and Stevenson, H. W. (1995). Response style and cross-cultural comparisons of rating scales among east asian and north american students. *Psychological Science*, 66(3):170–175.

Cheung, G. W. and Rensvold, R. B. (2000). Assessing extreme and acquiescence response sets in cross-cultural research using structural equations modeling. *Journal of Cross-Cultural Psychology*, 31(2):187–212.

Chun, K.-T., Campbell, J. B., and Yoo, J. H. (1974). Extreme response style in cross-cultural research: A reminder. *Journal of Cross-Cultural Psychology,* 5(4):465–480.

Clarke, I. (2000). Extreme response style in cross-cultural research: An empirical investigation. *Journal of Social Behavior and Personality*, 15(1):137–152.

Clarke, I. (2001). Extreme response in cross-cultural research. *International Marketing Review*, 18(3):301–324.

Converse, P. (1964). The nature of belief systems in mass publics. In Apter, D., editor, Ideology and Discontent. Free Press.

Crawford, S., Couper, M., and Lamias, M. (2001). Web surveys. perceptions of burden. *Social Science Computer Review*, 19:146—162.

Cronbach, L. J. (1946). Response set and test validity response set and test validity. *Educational and Psychological Measurement*, 6:475–494.

Dawes, C. T. and Fowler, J. (2009). Partisanship, voting, and the dopamine d2 receptor gene. *The Journal of Politics*, 71(3):1157–1171.

DeJong, M. G., Steenkamp, J.-B. E. M., Fox, J.-P., and Baumgartner, H. (2008). Using item response theory to measure extreme response style in marketing research: a global investigation. *Journal of Market Research*, 45:104–115.

Derks, E. M., Dolan, C. V., and Boomsma, D. I. (2006). A test of the equal environment assumption (eea) in multivariate twin studies. *Twin Research and Human Genetics*, 9(3):403–411.

Devlin, B., Daniels, M., and Roeder, K. (1997). The heritability of iq. *Nature*, 388:438–471.

Eaves, L. J. and Hatemi, P. K. (2008). Transmission of attitudes toward abortion and gay rights: Parental socialization or parental mate selection? *Behavior Genetics*, 38:247–256.

Eaves, L. J., Heath, A. C., Martin, N. G., Maes, H., Neale, M. C., Kendler, K., Kirk, K., and Corey, L. (1999). Comparing the biological and cultural inheritance of personality and social attitudes in the virginia 30000 study of twins and their relatives. *Twin Research*, 2:62–80.

Eysenck, H. J. and Prell, D. B. (1951). The inheritance of neuroticism: An experimental study. *Journal of Mental Science*, 97:441–465.

Falconer, D. S. (1960). Introduction to Quantitative Genetics. Edinburgh: Oliver and Boyd.

Feigon, S. A., Waldman, I. D., Levy, F., and Hay, D. A. (2001). Genetic and environmental influences on separation anxiety disorder symptoms and their moderation by age and sex. *Behavior Genetics*, 31(5):403–411.

Fowler, J., Baker, L., and Dawes, C. T. (2008). Genetic variation in political participation. *American Political Science Review*, 102(2):233–248.

Fowler, J. and Dawes, C. T. (2008). Two genes predict voter turnout. *The Journal of Politics*, 70(3):579–594.

Fowler, J. H. and Schreiber, D. M. (2008). Biology, politics, and the emerging science of human nature. *Science*, 322:912–914.

Freese, J. and Powell, B. (2003). Tilting at twindmills: Rethinking sociological responses to behavioral genetics. *Journal of Health and Social Behavior*, 44(2):130–135.

Fricke, B. G. (1956). Response set as a suppressor variable in the oais and mmpi. *Journal of Consulting Psychology*, 20:161–169.

Gilman, R., Huebner, E. S., Tian, L., Park, N., O'Byrne, J., Schiffand, M., Sverko, D., and Langknecht, H. (2008). Cross-national adolescent multidimensional life satisfaction reports: Analyses of mean scores and response style differences. *Journal of Youth and Adolescence*, 37:142–154.

Greenleaf, E. A. (1992). Measuring extreme response style. *Public Opinion Quarterly,* 56(3):328–351.

Groves, R. M., Fowler, F. J., Cooper, M. P., Lepkowski, J. M., Singer, E., and Tourangeau, R. (2004). Survey Methodology. Wiley Series in Survey Methodology. Wiley and Sons.

Guo, G. and Stearns, E. (2002). Social influence on the realization of genetic potential for intellectual development. *Social Forces*, 80(3):881–910.

Hammond, R. A. and Axelrod, R. (2006). Evolution of contingent altruism when cooperation is expensive. *Theoretical population biology*, 69(3):333–338.

Hannagan, R. and Hatemi, P. K. (2008). A comment on the genetics of political attitudes debate. *Perspectives on Politics*, 6(2):329–335.

Harden, K. P., Turkheimer, E., and Loehlin, J. C. (2007). Genotype by environment interaction in adolescents' cognitive aptitude. *Behavior Genetics*, 37:273–283.

Hatemi, P. K., Alford, J. R., Hibbing, J. R., Martin, N. G., and Eaves, L. J. (2009a). Is there a "party" in your genes? *Political Research Quarterly*, 62(3):584–600.

Hatemi, P. K., Alford, J. R., Hibbing, J. R., Medland, S. E., Keller, M. C., Martin, N. G., and Eaves, L. J. (2007a). Not by twins alone: Using extended family design to investigate the genetic basis of political beliefs. Presentated at the annaul meeting of the American Political Science Association, Chicago, IL, USA.

Hatemi, P. K., Funk, C. L., Medland, S. E., Maesa, H. M., Silberg, J. L., Martin, N. G., and Eaves, L. J. (2009b). Genetic and environmental transmission of political attitudes over a life time. *Journal of Politics*, 71:1141–1156.

Hatemi, P. K., Medland, S. E., and Eaves, L. J. (2009c). The genetic source of variance in the gender gap. *The Journal of Politics*, 71(1):262–276.

Hatemi, P. K., Morley, K. I., Medland, S. E., Heath, A. C., and Martin, N. G. (2007b). The genetics of voting: An australian twin study. *Behavior Genetics*, 37:435–448.

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.

Heath, A.C., Neale, M.C., Kessler, R.C., Eaves, L.J., and Kendler K.S. (1992). "Evidence for Genetic Influences on Personality From Self-Reports and Informant Ratings." *Journal of Personality and Social Psychology* 63: 85–96.

Heath, A. C., Nyholt D. R., Neuman, R., Madden P. A. F., Bucholz, K. K., Todd, R. D., Nelson, E. C., Montgomery, G. W., and Martin, N. G. (2003). Zygosity Diagnosis in the Absence of Genotypic Data: An Approach Using Latent Class Analysis. *Twin Research.* 6(1): 22–26

Hibbing, M. V., Mondak, J. J., and Anderson, M. R. (forthcoming). Personality and political discussion. *American Political Science Review*.

Hofstede, G. (1980). Culture's consequences: International studies of work-related values. Sage, Newbury Park, CA.

Hofstede, G. (2001). Culture's consequences: Comparing values, behaviors, institutions and organizations across nations. Sage, Thousand Oaks, CA.

Holbrook, A. L., Green, M. C., and Krosnick, J. A. (2003). Telephone versus face-to-face interviewing of national probability samples ith questionnaires: Comparisons of respondent satisficing and social desirability response bias. *Public Opinion Quarterly*, 67:79–125.

Horn, J. M., Plomin, R. and Rosenman, R. (1976). "Heritability of Personality Traits in Adult Male Twins." *Behavior Genetics* 6: 17-30.

Horwitz, A. V., Videon, T. M., Schmitz, M. F., and Davis, D. (2003a). Double vision: Reply to freese and powell. *Journal of Health and Social Behavior*, 44(2):136–141.

Horwitz, A. V., Videon, T. M., Schmitz, M. F., and Davis, D. (2003b). Rethinking twins and environments: Possible social sources for assumed genetic influences in twin research. *Journal of Health and Social Behavior*, 44(2):111–129.

Hui, C. H. and Triandis, H. C. (1985). The instability of response sets. *Public Opinion Quarterly*, 49(2):253–260.

Hui, C. H. and Triandis, H. C. (1989). Effects of culture and response format on extreme response styles. *Journal of Cross-Cultural Psychology*, 20(3):296–309.

Jackson, D. N. and Pacine, L. (1961). Response styles and academic achievement. *Educational and Psychological Measurement*, 21(4):1015–1028.

Jang, K. L., Mccrae, R. R., Livesley, W. J., Angleitner, A., and Riemann, R. (1998). Heritability of facet-level traits in a cross-cultural twin sample: Support for a hierarchical model of personality. *Journal of Personality and Social Psychology*, 74(6):1556–1565.

Javeline, D. (1999). Response effects in polite cultures: A test of acquiescence in kazakhstan. *The Public Opinion Quarterly*, 63(1):1–28.

Jinks, J. L. and Fulker, D. W. (1970). Comparison of the biometrical genetical, mava, and classical approaches to the analysis of human behavior. *Psychological Bulletin*, 73(5):311–349.

Johnson, T., Kulesa, P., Llc, I., Cho, Y. I., and Shavitt, S. (2005). The relation between culture and response styles: Evidence from 19 countries. *Journal of Cross-Cultural Psychology*, 36(2):264–277.

Jordan, L. A., Marcus, A. C., and Reeder, L. G. (1980). Response styles in telephone and household interviewing: A field experiment. *Public Opinion Quarterly*, 44(2):210–222.

Joseph, J. (2001). Separated twins and the genetics of personality differences: A critique. *American Journal of Psychology*, 114(1):1–30.

Keller, M. C., Medland, S. E., Duncan, L.E., Hatemi, P. K., Neale, M.C., Maes, H. H. M., and Eaves, L. J. (2009). Modeling extended twin family data I: description of the Cascade model. *Twin Research and Human Genetics*, 12(1):8-18

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., Karkowski, L. M., Neale, M. C., and Prescott, C. A. (2000). Illicit psychoactive substance use, heavy use, abuse, and dependence in a us population-based sample of male twins. *Archives of General Psychiatry*, 57:261–269.

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

Kiesler, S. and Sproull, L. S. (1986). Response effects in the electronic survey. *The Public Opinion Quarterly*, 50(3):402–413.

Krosnick, J. A. (1991). Response strategies of coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology*, 5:213–236.

Krosnick, J. A. (1999). Survey research. *Annual Review of Psychology*, 50:537–67.

Kuethe, J. (1960). Acquiescent response set and the psychasthenia scale: an analysis via the aussage experiment. *Journal of Abnormal and Social Psychology*, 61:319–22.

Lamm, B. and Keller, H. (2007). Understanding cultural models of parenting: The role of intracultural variation and response style. *Journal of Cross-Cultural Psychology*, 38(1):50–57.

Landsberger, H. A. and Saavedra, A. (1967). Response set in developing countries. *The Public Opinion Quarterly*, 31(2):214–229.

Layton, R. I. (1982). Personality Assesment. Wiley, New York.

Lenski, G. E. and Leggett, J. C. (1960). Caste, class, and deference in the research interview. *The American Journal of Sociology*, 65(5):463–467.

Lentz, T. F. (1938). Acquiescence as a factor in the measurement of personality. *Psychological Bulletin*, 35:659.

Leung, K. and Bond, M. H. (1989). On the empirical identification of dimensions for cross-cultural comparisons. *Journal of Cross-Cultural Psychology*, 20(2):133–151.

Lewis, N. A. and Taylor, J. A. (1955). Anxiety and extreme response preferences. *Educational and Psychological Measurement*, 15:111–116.

Light, C. S. and Zax, M. (1965). Relationship of age, sex, and intelligence level to extreme response style. *Journal of Personality and Social Psychology*, 2(6):907–909.

Loehlin, J. C. and Nichols, R. C. (1976). Heredity, environment, and personality: A study of 850 sets of twins. Austin, TX: University of Texas Press.

Lorge, I. (1937). Gen-like: Halo or reality? *Psychological Bulletin*, 34:545 –546.

Madsen, D. (1985). A biochemical property relating to power seeking in humans. *The American Political Science Review*, 79(2):448–457.

Madsen, D. (1986). Power seekers are different: Further biochemical evidence. *The American Political Science Review*, 80(1):261–269.

Maes, H. H., Neale, M. C., Kendler, K. S., Martin, N. G., Heath, A. C., and Eaves, L. J. (2006). Genetic and cultural transmission of smoking initiation: An extended twin kinship model. *Behavior Genetics*, 36(6):795–808.

Maes, H. H. M., Neale, M. C., and Eaves, L. J. (1997). Genetic and environmental factors in relative body weight and human adiposity. *Behavior Genetics*, 27:325–351.

Marin, G., Gamba, R. J., and Marin, B. V. (1992). Extreme response style and acquiescence among Hispanics: The role of acculturation and education. *Journal of Cross-Cultural Psychology*, 23(4):498–509.

McCrae, R. R., and P. T. Costa, Jr. (2003) Personality in Adulthood: A Five-Factor Theory Perspective, 2nd ed. New York: The Guilford Press.

Medland, S. E. and Hatemi, P. K. (2009). Political science, behavior genetics and twin studies: a methodological primer. *Political Analysis*, 17:191–214.

Mitchell, K. S., Mazzeo, S. E., Bulik, C. M., Aggen, S. H., Kendler, K. S., and Neale, M. C. (2007a). An investigation of a measure of twins' equal environments*. Twin Research and Human Genetics*, 10(6):840–847.

Mitchell, K. S., Mazzeo, S. E., Schmitt, J. E., Bulik, C. M., and Neale, M. C. (2007b). Assessing violations of the equal environments assumption with a finite mixture distribution model. *Behavior Genetics*, 37(6):777.

Morris, M. W., Williams, K. Y., Leung, K., Larrick, R., Mendoza, M. T., Bhatnagar, D., Li, J., Kondo, M., Luo, J.-L., and Hu, J.-C. (1998). Conflict management style: Accounting for cross-national differences. *Journal of International Business Studies*, 29:729–748.

Muthen, B., Asparouhov, T., and Rebollo, I. (2006). Advances in behavioral genetics modeling using mplus: Applications of factor mixture modeling to twin data. *Twin Research and Human Genetics,* 9:313–324.

Muthen, L. K. and Muthen, B. O. (2007). Mplus Version 5, Statistical Analysis With Latent Variables. User's Guide.

Narayan, S. and Krosnick, J. A. (1996). Education moderates some response effects in attitude measurement. *Public Opinion Quarterly*, 60:58–88.

Neale, M. C. and Cardon, L. R. (1992). Methodology for genetic studies of twins and families. Dordrecht: Kluwer Academic Publishers BV.

Neale M.C. (2003). A finite mixture distribution model for data collected from twins. *Twin Research* 6(3):235-239

Nemeth, A. and Takacs, K. (2007). The evolution of altruism in spatially structured populations. *Journal of Artificial Societies and Social Simulation*, 10(3):4.

Ostini, R. and Nering, M. L. (2005). Polytomous Item Response Theory Models, volume 144 of Quantitative Applications in the Social Sciences. Sage, Thousand Oaks, CA.

Oxley, D., Smith, K., Alford, J., Hibbing, M., Miller, J., Scalora, M., Hatemi, P., and Hibbing, J. (2008). Political attitudes vary with physiological traits. *Science*, 321:1667–1701.

Pedersen, N. L., Plomin, R., McClearn, G.E., Friberg, L. (1988) "Neuroticism, Extraversion, and Related Traits in Adult Twins Reared Apart and Reared Together." *Journal of Personality and Social Psychology* 55: 950-7.

Plomin, R., Pedersen, N. L., Lichtenstein, P., and McClearn, G. E. (1994). Variability and stability in cognitive abilities are largely genetic later in life. *Behavior Genetics*, 24(3):207–215.

Plomin, R., Willerman, L., and Loehlin, J. C. (1976). Resemblance in appearance and the equal environments assumption in twin studies of personality trait. *Behavior Genetics*, 6:43–52.

Posthuma, D. and Boomsma, D. I. (2005). Mx scripts library: Structural equation modeling scripts for twin and family data. *Behavior Genetics,* 35(4):499–505.

Purcell, S. (2002). Variance components models for gene–environment interaction in twin analysis. *Twin Research*, 5(6):554–571.

Richardson, K. and Norgate, S. (2005). The equal environments assumption of classical twin studies may not hold. *British Journal of Educational Psychology*, 75(3):339–350.

Riemann, R., Angleitner, A., Strelau, J. (1997) Genetic and Environmental Influences on Personality: A Study of Twins Reared Together Using the Self- and Peer Report NEO-FFI Scales. *Journal of Personality* 65(3):449-475

Rietveld, M., Dolan, C., van Baal, G., and Boomsma, D. (2003). A twin study of differentiation of cognitive abilities in childhood. *Behavior Genetics*, 33:367–381.

Rose, R. J., Kaprio, J., Williams, C. J., Viken, R., and Obremski, K. (1990). Social contact and sibling similarity: Facts, issues, and red herrings. *Behavior Genetics*, 20:763–778.

Rose, R. J., Koskenvuo, M., Kaprio, J., Sarna, 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.

Ross, C. E. and Mirowsky, J. (1996). Socially-desirable response and acquiescence in a cross-cultural survey of mental health. *Journal of Health and Social Behavior*, 25(2):189–197.

Roster, C., Albaum, G., and Rogers, R. (2006). Can cross-national/cultural studies presume etic equivalency in respondents' use of extreme categories of likert rating scales? *International Journal of Market Research*, 48(6):741–759.

Rushton, J. P., Fulker, D. W., Neale, M. C., Nias, D. K. B., and Eysenck, H. J. (1986). Altruism and aggression: The heritability of individual differences. *Journal of Personality and Social Psychology*, 50:1192–1198.

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

Schwartz, S. H. (1994). Beyond individualism/collectivism: New cultural dimensions of values. In Kim, U., Triandis, H. C., Kagitcibasi, C., Choi, S.-C., and Yoon, G., editors, Individualism and collectivism: Theory, method and applications, pages 85–119. Sage, Thousand Oaks, CA.

Settle, J. E., Dawes, C. T., and Fowler, J. H. (2009). The heritability of partisan attachment. *Political Research Quarterly*, 62(3):601–613.

Shulman, A. (1973). A comparison of two scales on extremity response bias. *Public Opinion Quarterly*, 37(3):407–412.

Sigelman, L. (2006). Report of the editor of the american political science review, 2004-05. *PS: Political Science & Politics*, 40:171–173.

Smith, P. B. (2004). Acquiescent response bias as an aspect of cultural communication style. *Journal of Cross-Cultural Psychology*, 35(1):50–61.

Smith, P. B., Peterson, M. F., and Schwartz, S. H. (2002). Cultural values, sources of guidance, and their relevance to managerial behavior. *Journal of Cross-Cultural Psychology*, 33(2):188–208.

Spector, P. E. (1987). Method variance as an artifact in self-reported affect and perceptions at work: Myth or significant problem*? Journal of Applied Psychology*, 72(3):438–443.

Stening, B. W. and Everett, J. E. (1984). Response styles in a cross-cultural managerial study. *Journal of Social Psychology*, 122:151–156.

Sudman, S. S., Bradburn, N. M., and Schwarz, N. (1996). Thinking About Answers. Jossey-Bass.

Tellegen, A., Lykken, D. T., Bouchard, T. J. Jr., Wilcox, K. J., Segal, N. L., Rich, S. (1988). "Personality Similarity in Twins Reared Apart and Together." *Journal of Personality and Social Psychology* 54: 1031-9.

Truett, K. R., Eaves, L. J., Walters, E. E., Heath, A. C., Hewitt, J. K., Meyer, J. M., Silberg, J., Neale, M. C., Martin, N. G., and Kendler, K. S. (1994). A model system for analysis of family resemblance in extended kinships of twins. *Behavior Genetics,* 24(1):35–49.

Tupes, E. C., and R. E. Christal. 1958. Stability of Personality Trait Rating Factors Obtained under Diverse Conditions. USAF WADC Technical Note No. 58-61. Lackland Air Force Base, TX: U.S. Air Force.

Tupes, E. C., and R. E. Christal. 1961. Recurrent Personality Factors Based on Trait Ratings USAF ASD Technical Report No. 61-97. Lackland Air Force Base. TX: U.S. Air Force.

van Beijsterveldt, C. E. and van Baal, G. C. (2002). Twin and family studies of the human electroencephalogram: a review and a meta-analysis. *Biological Psychology*, (61):111–138.

van de Vijver, F. J. R. and Leung, K. (1997). Methods and data analysis for cross-cultural research. Sage, Newbury Park, CA.

van Herk, H., Poortinga, Y. H., and Verhallen, T. M. M. (2004). Response styles in rating scales: Evidence of method bias in data from six eu countries. *Journal of Cross-Cultural Psychology*, 35(3):346–360.

Visscher, P. M., Medland, S. E., Ferreira, M. A. R., Morley, K. I., Zhu, G., Cornes, B. K., Montgomery, G. W., and Martin, N. G. (2006). Assumption-free estimation of heritability from genome-wide identity-by-descent sharing between full siblings. *PLoS Genet*, 2:e41.

Watkins, D. and Cheung, S. (1995). Culture, gender, and response bias: An analysis of responses to the self-description questionnaire. *Journal of Cross-Cultural Psychology*, 26(5):490–504.

Xian, H., Scherrer, J. F., Eisen, S. A., True, W. R., Heath, A. C., Goldberg, J., Lyons, M. J., and Tsuang, M. T. (2000). Self-reported zygosity and the equal-environments assumption for psychiatric disorders in the vietnam era twin registry. *Behavior Genetics*, 30:303–310.

Appendix A

[Note, this information is directly taken from MIDUS study documentations]

We cannot compute an exact response rate for MIDUS because we only wanted to interview about half of people we contacted, and it is only the latter who should be in the denominator of the response rate. We have no way of knowing how many of the refusers would have been selected for interview, which means that we cannot compute the denominator exactly. However, it is possible to make an estimate of the number of people in the denominator and, based on this, an estimate of the response rate. This estimate, the calculation of which is described below, is 70.0% for the telephone interview, 86.8% for the completion of the main questionnaire among the telephone respondents, and 60.8% for the overall response rate (0.700 x 0.868).

Comparison of the MIDUS I sample with the Current Population Survey (CPS, 1995) revealed that the sample under-represented those with a high school education or less and African Americans. Alternatively, it over-represented older males (by intention, to facilitate gender comparisons by age). The representation by gender and marital status was close to the CPS.

Of the original 7,190 individuals from whom data were collected at MIDUS 1 [including the general survey, the twin, the sibling and the urban oversamples], the University of Wisconsin Survey Center (UWSC) was able to collect data from 4,975 cases 10 years later. However, 842 original participants refused to participate and 1,334 were not successfully recontacted. This last group included cases for whom mortality was confirmed or suspected, those who were too ill to take the survey, and all other cases that the UWSC was unable to contact.

Appendix B: Questions measuring Acquiescence.

(Response categories: 1 strongly agree, 2 somewhat agree, 3 agree a little, 4 don't know, 5 disagree a little, 6 somewhat disagree, 7 strongly disagree)

<...>
A7 Please indicate how much you agree or disagree with the following statements.
<...>
e. When I am sick, getting better is in the doctor's hands
f. It is difficult for me to get good medical care

<...>
F1 Please indicate how strongly you agree or disagree with each of the following statements.
<...>
s. There is little I can do to change the important things in my life
t. I often feel helpless in dealing with the problems of life
w. What happens in my life is often beyond my control
y. There are many things that interfere with what I want to do
<…>

<...>
K17. Please indicate how strongly you agree or disagree with each of the following statements.
a. The world is too complex for me.
b. I don't feel I belong to anything I'd call a community
c. People who do a favor expect nothing in return
<...>
e. The world is becoming a better place for everyone
f. I feel close to other people in my community
g. My daily activities do not create anything worthwhile for my community
h. I cannot make sense of what's going on in the world
i. Society has stopped making progress
j. People do not care about other people's problems
k. My community is a source of comfort
l. I find it easy to predict what will happen next in society
m. Society isn't improving for people like me
<...>

<…>

Appendix C: Questions measuring Extreme Response Style.

<…>
A15. During the past 30 days, how much of the time did you feel...
1,All of the time 2,Most of the time 3,Some of the time 4,A little of the time 5,None of the time
<…>
f. ...full of life?
<…>

<…>
F1. Please indicate how strongly you agree or disagree with each of the following statements.
1 strongly agree, 2 somewhat agree, 3 agree a little, 4 don't know, 5 disagree a little, 6 somewhat disagree, 7 strongly disagree
<…>
y. There are many things that interfere with what I want to do
<…>

<…>
F6. The next few questions are about the way you decide what you want out of life and how you go about trying to achieve your goals. For each situation below, two different strategies are listed. Please indicate whether your own strategy is more like the one listed in column A or the one listed in column B. While the way you do things may be different depending on the particular goal, and may include parts of both strategies, please circle the answer that is true for you overall.

MORE LIKE A 1, A Lot - 2, Some - 3, A Little - 4, A Little - 5, Some - 6, A Lot MORE LIKE B

STRATEGY A	STRATEGY B

F6a. When choosing my goals...

I prefer to choose one or two important goals and really focus on achieving them.	I prefer not to limit myself -- I keep my options open so I can take advantage of anything that comes up.

<…>
F6c. If I don't seem to have a particular skill or resource that I need to reach my goal...

I look for other things I could do to reach my goal -- to make up for what I don't have or can't do.	I keep trying my best, and if that doesn't work, I think again about whether that goal is right for me.

<…>

<…>
K17. Please indicate how strongly you agree or disagree with each of the following statements.
1 strongly agree, 2 somewhat agree, 3 agree a little, 4 don’t know, 5 disagree a little, 6 somewhat disagree, 7 strongly disagree
<…>
c. People who do a favor expect nothing in return
<…>
j. People do not care about other people's problems
k. My community is a source of comfort
<…>

<…>
Q5. Using a 0 to 10 scale where 0 means "no thought or effort" and 10 means "very much thought and effort," how much thought and effort do you put into the sexual aspect of your life these days?
None 00 01 02 03 04 05 06 07 08 09 10 Very Much

Appendix D: ACE Model Fit Compared to the Respective Baseline Models

	-2LL	Chi-Square	df	p-value
Extreme Response Style				
Phenotypic Saturated Model	16368.972			
ACE Model	16376.758	7.786	6	0.25
Extreme Response Style w/ Education				
Phenotypic Saturated Model	16001.668			
ACE Model	16008.9	7.232	6	0.30
Extreme Response Style w/ Income				
Phenotypic Saturated Model	22156.824			
ACE Model	22164.146	7.322	6	0.29
Acquiescence 1				
Phenotypic Saturated Model	19213.212			
ACE Model	19220.166	6.954	10	0.73
Acquiescence 1 w/ Education				
Phenotypic Saturated Model	24960.928			
ACE Model	24966.234	5.306	10	0.87
Acquiescence 1 w/ Income				
Phenotypic Saturated Model	24778.632			
ACE Model	24785.604	6.972	10	0.73
Acquiescence 2				
Phenotypic Saturated Model	20702.882			
ACE Model	20710.726	7.844	10	0.64
Acquiescence 2 w/ Education				
Phenotypic Saturated Model	26394.462			
ACE Model	26399.71	5.248	10	0.87
Acquiescence 2 w/ Income				
Phenotypic Saturated Model	26243.546			
ACE Model	26251.856	8.31	10	0.60
Acquiescence 3				
Phenotypic Saturated Model	21682.208			
ACE Model	21689.9	7.692	10	0.66
Acquiescence 3 w/ Education				
Phenotypic Saturated Model	27366.442			
ACE Model	27372.732	6.29	10	0.79
Acquiescence 3 w/ Income				
Phenotypic Saturated Model	27217.57			
ACE Model	27225.242	7.672	10	0.66

Note: Degrees of freedom for Extreme Response Style is less then for Acquiescence. The Extreme Response Style model has fewer estimates because the IRT model used to produce the latent scores restricts their variance to 1. For this reason total variance of the phenotype is not an estimate and Unique Environment is calculates as 1 – Heritability - Common Environment.