Childhood Autism and Assortative Mating

Hays Golden*

University of Chicago

November 20, 2012

Abstract

Diagnosed rates of autism spectrum disorders have grown tremendously over the last few decades. I find that assortative mating may have meaningfully contributed to the rise. I develop a general model of genes and assortative mating which shows that small changes in sorting could have large impacts on the extremes of genetic distributions. I apply my theory to autism, which I model as the extreme right tail of a genetic formal thinking ability distribution (systemizing). Using large sample data from the Centers for Disease Control and Prevention, I find strong support for theories that autism is connected to systemizing. My mating model shows that increases in the returns to systemizing, particularly for women, can contribute significantly to rising autism rates. I provide evidence that mating on systemizing has actually shifted, and conclude with a rough calculation suggesting that despite the increase in autism, increased sorting on systemizing has been socially beneficial.

*This paper benefitted greatly from conversations with Scott Kominers, Gary Becker, Steven Levitt, Derek Neal, Kevin Murphy, Roger Myerson, Julian Reif, and others. I would also like to thank seminar participants at the University of Chicago Micro Lunch and the Labor Micro Working Group. I’m especially grateful to Jon Baio at the CDC for his support and comments. Aggregate counts of children aged eight years identified with select developmental disabilities were furnished by the U.S. Centers for Disease Control and Prevention (CDC) Metropolitan Atlanta Developmental Disabilities Surveillance Program (MADDS)., a multiple source, population-based disability monitoring system covering five counties in Atlanta, Georgia. The findings and conclusions in this article are those of the author and do not necessarily represent the official position of the Centers for Disease Control and Prevention.
1 Introduction

Positive assortative mating, or the tendency of likes to mate with likes, has long been observed across many traits, and even across many species.\textsuperscript{1} At the same time, assortative mating has become much easier over the last century. An increasingly urban population has created deeper local marriage markets. Higher rates of college attendance, especially among women, have given people an early-in-life opportunity to meet similar potential mates: if you marry your high school sweetheart, you picked among those people who happened to be born in your town, but if you marry your college sweetheart, you have picked among people who chose the same college that you did. More recently, the internet has made it easier to find people with similar interests or traits, and increasingly popular dating sites, like Match.com, have even automated the process.

While economists have largely focused on the benefits of positive assortative mating that arise from, for instance, complementarities in the household production function, one less understood impact of assortative mating is the strong effect it has on the extremes of distributions. Positive assortative mating on a trait increases the variance of that trait, and the increase in variance causes large relative growth at the extremes of the distribution.\textsuperscript{2} I explore a particular example of this, where the growing returns to mathematical ability may have contributed significantly to the increased prevalence of autism

\textsuperscript{1}Thiessen and Gregg (1980)
\textsuperscript{2}I discuss this more formally in section 3.5.
spectrum disorders.

Diagnosed rates of autism spectrum disorders (ASDs) have grown tremendously over the last few decades, and a number of possible explanations have emerged. One of the more interesting possibilities is that ASDs are related to a genetic trait called systemizing, which governs how much our brains are wired for thinking about formal systems.\(^3\) High levels of systemizing may give ability in pursuits like mathematics and computer languages. In this view, ASDs occur when someone has too much systemizing: when their brains are so wired for formal systems that they begin to have trouble dealing with systems that do not follow strict rules.

Since the market returns to abilities driven by systemizing have grown over the last few decades, and market opportunities to use these abilities seem to be more open to women than they once were, we may expect increased assortative mating on systemizing. More assortative mating on a genetic trait causes the variance of that trait to increase in the population, so we would expect to see more children at the right tail of the systemizing distribution, and so more children with ASDs.

Since ASDs are relatively rare (60 cases per 10,000 for ASDs overall, and between 10 and 20 cases per 10,000 for autism\(^4\)), increases in the variance will have a large impact on the portion of the population with ASDs. Specifically, I calculate in section 3.5 that the portion of people with ASDs should grow seven times as quickly as the standard deviation, and the portion of people with autism should grow ten times as quickly as the standard deviation. This kind

\(^3\)Baron-Cohen (2006)  
\(^4\)Newschaffer et al. (2007)
of rapid growth really does happen, as we can see from the example of height. In section 3.5, I show that the standard deviation of the height distribution in the US grew 10% in the last fifty years. Being shorter than 4'10" or taller than 6'2" are both about as rare as having an ASD, and my model predicts that each group should have had relative growth of about 65%. Accounting for the change in the mean, the share of those under 4'10" actually increased 92%, and the share of those over 6'2" actually increased 73%, both even higher than the already very high prediction. If the systemizing distribution had its standard deviation grow as much as the height distribution actually did, that alone would explain a doubling in autism rates. Since mating on systemizing may have increased at a greater rate than mating on height, this could explain even greater rates of growth if the standard deviation grew more.

I explore this possibility by first extending the evidence for a connection between ASDs and systemizing in a large sample, general population dataset in section 2. In section 3, I present a general model of genetic distributions and assortative mating. The model comes from taking a standard model in quantitative genetics and giving microfoundations to the mating process, the result has similarities to Becker and Tomes (1979) and Kremer (1997). In section 4, I provide some evidence on how mating has changed. Finally, in section 5, I do a rough calculation which suggests that increased assortative mating on systemizing has probably been positive, despite the increase in autism rates.
2 Connecting autism to mathematical ability

Simon Baron-Cohen, a psychologist at Cambridge University, argues in a 2006 paper that human brains have a capacity to systemize, to detect relationships in information, and that autism spectrum disorders occur when the brain is too wired for systemizing.\textsuperscript{5} He outlines different levels of systemizing, where higher levels initially give greater ability to understand abstract systems. However, as the brain becomes more wired for systemizing, you may see impairment because people become increasingly unable to deal with systems that do not have a lawful structure, such as social interactions and verbal communication. Further, people at a very high systemizing level may become intensely interested in lawful, predictable things like train schedules.\textsuperscript{6} Those three features, social impairment, communications impairment, and restricted and repetitive behaviors and interests, are the three components of the diagnostic definition of autism.\textsuperscript{7}

There is very strong evidence that autism has a substantial genetic component. Twin studies typically indicate a heritability of at least .9, meaning that at least 90\% of the variation in autism status is due to variation in genes.\textsuperscript{8}

Baron-Cohen and co-authors have conducted a number of suggestive stud-

\textsuperscript{5}Baron-Cohen (2006)
\textsuperscript{6}Baron-Cohen (2006)
\textsuperscript{7}DSM-IV-TR (2000)
\textsuperscript{8}Freitag (2006) provides a review of the literature. Twin studies compare the correlation in a trait between identical and fraternal twins to measure how much of the variation in a trait is due to genes. Early studies relied on statistical rules of thumb, but the results have held up in recent studies using more rigorous methods. Additionally, twin studies underestimate the contributions of genes in the presence of positive assortative mating because fraternal twins will be more genetically similar than the estimation methodology assumes.
ies which provide some empirical support for the link between ASD and systemizing. These include a study showing a higher rate of prevalence of autism among math majors at Cambridge University and their relatives, compared to control students, even controlling for sex,\textsuperscript{9} one showing a higher prevalence of engineers among the relatives of children with autism,\textsuperscript{10} as well a study showing that parents of children with Asperger’s Syndrome, a mild ASD, show more systemizing traits on tests, compared to a control sample.\textsuperscript{11}

Baron-Cohen also led a study (Roelfsema et al., 2011) which looked at the autism prevalence in three regions of the Netherlands: Eindhoven, Haarlem, and Utrecht. Eindhoven has 30\% of its population employed in information technology, compared with 16\% and 17\% for the two other regions. As predicted, Eindhoven has a much higher prevalence of childhood autism.

Building off of these interesting results, I want to test the connection between systemizing and autism in a large, general population sample. One source of data comes from the Metropolitan Atlanta Developmental Disabilities Surveillance Program (MADDSP), which is run by the Centers for Disease Control and Prevention (CDC). MADDSP tries to get a complete count of 8 year olds in a five county area of Metro Atlanta who have an ASD, and also tracks certain other developmental disabilities. Potential cases are identified through the records of many sources including hospitals, clinics, specialty providers, and all of the area’s public schools.\textsuperscript{12} Clinicians review the behavioral information collected and make a determination, spending an average of

\textsuperscript{9}Baron-Cohen et al. (2007)  
\textsuperscript{10}Baron-Cohen et al. (1997)  
\textsuperscript{11}Baron-Cohen and Hammer (1997)  
\textsuperscript{12}Rice et al. (2007)
47 minutes on each child who does not have a previous clinical diagnosis.\textsuperscript{13} In 2006, 2.4\% of 8 year olds living in the study area had their records reviewed by a clinician as part of the study.

I have data from five study years (2000, 2002, 2004, 2006, and 2008) which give ASD case counts by Census block group. Among the 1337 block groups, there are 2098 cases, pooling across the five years. Because I am pooling the five cohorts, I use the population of children 17 and under in each block group from the 2000 Census, multiplied by $\frac{5}{18}$, as the denominator for computing prevalence. Those 1337 block groups contain over 1\% of all children in the US.

This dataset allows me to overcome several of the challenges faced by Roelfsema et al., since I have 1337 observations to their 3, a uniform diagnostic methodology across the regions, and complete response where they had only partial response. The larger number of observations also allows me to consider more covariates.

Systemizing among the adults is more difficult to measure, but is proxied for by occupation. Summary File 3 of the Census gives occupations for employed adults in a one in six sample of households in each block group, with the occupations grouped into 33 categories.

To determine which occupations are more mathematically intensive, I use the Department of Labor’s O*Net Project, which has measurements of the importance of different skills in different occupations. I use the measure of mathematics importance, which ranges from 1-100, and comes from a com-

\textsuperscript{13}Van Naarden Braun et al. (2007)
bination of worker surveys and expert evaluation.\textsuperscript{14} O*NET provides scores for individual occupations, so I use Occupational Employment Statistics data to weight the occupation scores by their frequency and get the category averages. The categories averages allow me to get a systemizing score for each block group which I call the average math importance.

Durkin, et al (2010) showed a relationship between measures of socioeconomic status and autism prevalence in a dataset that included the MADDSP data I am using.\textsuperscript{15} I include, as controls, the same measures that they used, median household income, the poverty rate, and the percent of adults with a bachelor’s degree.\textsuperscript{16} As a placebo, I also regress all of these measures on cerebral palsy (CP) prevalence as measured by MADDSP using the same methodology. Like ASDs, impairment from CP can be mild, moderate, or severe. There is no reason to believe that systemizing is related to CP, so this can help us to see which variables are related to detection, and which are related to true prevalence.

\textsuperscript{14}O*NET 15.0 Database [Database]. U.S. Department of Labor, National Center for O*NET Development. Available at: http://www.onetcenter.org/database.html.

\textsuperscript{15}Durkin et al. (2010). They have the same Metro Atlanta data I do, but have data from other sites around the country as well. I am currently working on getting data from other sites as part of a collaborative project with researchers at those sites.

\textsuperscript{16}All measures come from the 2000 Census and all are also provided at the block group level. While 2010 data are becoming available, the block group boundaries have changed.
<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Avg Math Importance</td>
<td>0.264***</td>
<td>0.0164</td>
<td>0.289***</td>
<td>0.0634</td>
</tr>
<tr>
<td></td>
<td>(0.0237)</td>
<td>(0.0260)</td>
<td>(0.0552)</td>
<td>(0.0596)</td>
</tr>
<tr>
<td>Poverty Rate</td>
<td></td>
<td>-0.128***</td>
<td>-0.110**</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.0307)</td>
<td>(0.0371)</td>
<td></td>
</tr>
<tr>
<td>% of Adults with BA</td>
<td></td>
<td>-0.138*</td>
<td>-0.101</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.0627)</td>
<td>(0.0625)</td>
<td></td>
</tr>
<tr>
<td>Med. Household Income</td>
<td></td>
<td>0.00298</td>
<td>-0.0496</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.0483)</td>
<td>(0.0501)</td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>1,333</td>
<td>1,333</td>
<td>1,332</td>
<td>1,332</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.070</td>
<td>0.000</td>
<td>0.084</td>
<td>0.011</td>
</tr>
</tbody>
</table>

All variables are in standard deviation units

Observations are weighted by the number of child residents

Robust standard errors in parentheses

*** p<0.001, ** p<0.01, * p<0.05

The results for socio-economic status variables are similar for ASD and CP, but there is a large and significant effect of the mathematics importance scores on ASD, and no significant effect on CP. That is strong support for the connection between ASDs and systemizing, and is all the stronger considering that I am using block-group level variables to proxy for individual parental traits.
3 Assortative Mating and Genetic Distributions

My explanation of the rise in autism is increased assortative mating on systemizing, so we need an explanation of why assortative mating on systemizing increased, and what impact that would have. To construct a formal model of the effects of endogenous assortative mating on genetic distributions, I build upon a textbook model in quantitative genetics. Bulmer (1980) presents a model in which a single continuous trait (e.g. systemizing) is normally distributed in the population. Parents are drawn (in a possibly correlated way) from the population distribution and make equal genetic contributions to children, and children get a random environmental shock. Given assumptions, Bulmer can characterize the evolution of the distribution, and find the stationary distribution.

I extend Bulmer’s model by generalizing it to two traits and creating an explicit, though simple, marriage market, so the primitives of my model are distributions and returns to abilities instead of distributions and spousal correlations.

The model below has some similarities to Becker and Tomes (1979) and Kremer (1997), both of which look, at least in part, at the long run variance of a trait driven by intergenerational processes. However, my model has some key differences. Becker and Tomes look at optimizing dynasties and derive the long-run distribution of income from the choices of dynasties. But they have no marriage market and asexual reproduction, so the population distribution
Kremer models a child’s educational attainment to depend on the average attainment of the child’s parents and the average attainment of neighbors. He then derives the long-run variance of education, but exogenously assumes the correlation in education among spouses.

In my model, agents have two traits, systemizing ability ($S$) and general ability ($G$). Each trait comes partially from genes and partially from environmental factors.

$$S = S_g + S_E$$ (1)

$$G = G_g + G_E$$ (2)

Genetic contributions are normally distributed.

$$
\begin{pmatrix} G_g \\ S_g \end{pmatrix} \sim N \left[ \begin{pmatrix} \mu_G \\ \mu_S \end{pmatrix}, \begin{pmatrix} \sigma^2_{G,g} & \text{Cov}_{G,g} \\ \text{Cov}_{G,s} & \sigma^2_{S,s} \end{pmatrix} \right] (3)
$$

Environmental contributions are also normally distributed.

$$
\begin{pmatrix} G_E \\ S_E \end{pmatrix} \sim N \left[ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2_{G,E} & \text{Cov}_{G,E} \\ \text{Cov}_{E,E} & \sigma^2_{S,s} \end{pmatrix} \right] (4)
$$

Becker and Tomes note that marriage could be introduced conceptually in their model by assuming perfect assortative mating so that each person mates with a clone of themselves. In my model, that would correspond to assuming no environmental inputs. Although I will have perfect assortative mating on some quantity ($W$), it is not perfect assortative mating on the variable of interest ($G_g$) as it would be in that interpretation of Becker and Tomes.
Agents have a wage which comes from their level of each ability, and the returns to systemizing ability may differ by sex.

\[ W_m = G \cdot W_G + S \cdot W_S \]  
\[ = (G_\gamma + G_E)W_G + (S_\gamma + S_E)W_S \]  
\[ W_f = (G_\gamma + G_E)W_G + (S_\gamma + S_E)\alpha W_S \]  
\[ 0 \leq \alpha \leq 1 \]

I assume that the marital surplus function is supermodular in each \( W \), so I will have strict positive assortative mating on the values of \( W \).\(^{18}\) Each couple has one son and one daughter, whose genes are averages of their parents, plus a shock.

\[ G_\gamma = \frac{1}{2} G_{\gamma,Father} + \frac{1}{2} G_{\gamma,Mother} + \epsilon_G \]  
\[ S_\gamma = \frac{1}{2} S_{\gamma,Father} + \frac{1}{2} S_{\gamma,Mother} + \epsilon_S \]

\(^{18}\)We can think of \( W \) as a labor market wage, with positive assortative mating coming from complementarity in the household production function, or we can think of \( W_G \) and \( W_S \) as the marriage market returns to each ability. This approach will be discussed later in the section.

I am not explicitly including investment in abilities, but we can think of the returns to each ability as reflecting optimal investment given genetic and environmental shocks.
Where

\[
\begin{pmatrix}
\epsilon_G \\
\epsilon_S
\end{pmatrix} \sim N\left[
\begin{pmatrix}
0 \\
0
\end{pmatrix},
\begin{pmatrix}
\text{Var}(\epsilon_G) & \text{Cov}(\epsilon_G, \epsilon_S) \\
\text{Cov}(\epsilon_G, \epsilon_S) & \text{Var}(\epsilon_S)
\end{pmatrix}\right]
\]

(9)

Each child gets an independent joint draw of $\epsilon_G$ and $\epsilon_S$. Children draw from the environmental distribution to determine their final levels of $G$ and $S$. Members of this generation then marry and give birth to another generation.

3.1 Equilibrium

A stationary equilibrium will have a stationary genetic distribution. The means are stationary by construction since the dynamics consist just of averaging and mean zero shocks. For the genetic covariance matrix to be stationary, we will need:
\[
\text{Var}(G) = \text{Var} \left( \frac{1}{2} G_{\gamma, \text{Father}} + \frac{1}{2} G_{\gamma, \text{Mother}} + \epsilon_G \right) \quad (10)
\]

\[
= \frac{1}{2} \text{Var}(G) + \frac{1}{2} \text{Cov}(G_{\gamma, \text{Father}}, G_{\gamma, \text{Mother}}) + \text{Var}(\epsilon_G)
\]

\[
\text{Var}(S) = \text{Var} \left( \frac{1}{2} S_{\gamma, \text{Father}} + \frac{1}{2} S_{\gamma, \text{Mother}} + \epsilon_S \right) \quad (11)
\]

\[
= \frac{1}{2} \text{Var}(S) + \frac{1}{2} \text{Cov}(S_{\gamma, \text{Father}}, S_{\gamma, \text{Mother}}) + \text{Var}(\epsilon_S)
\]

\[
\text{Cov}(G, S) = \text{Cov} \left( \frac{1}{2} G_{\gamma, \text{Father}} + \frac{1}{2} G_{\gamma, \text{Mother}} + \epsilon_G, \frac{1}{2} S_{\gamma, \text{Father}} + \frac{1}{2} S_{\gamma, \text{Mother}} + \epsilon_S \right) \quad (12)
\]

\[
= \frac{1}{2} \text{Cov}(G, S) + \frac{1}{4} \text{Cov}(G_{\gamma, \text{Father}}, S_{\gamma, \text{Mother}})
\]

\[
+ \frac{1}{4} \text{Cov}(G_{\gamma, \text{Mother}}, S_{\gamma, \text{Father}}) + \text{Cov}(\epsilon_G, \epsilon_S)
\]

\[
= \frac{1}{2} \text{Cov}(G_{\gamma, \text{Father}}, S_{\gamma, \text{Mother}}) + \frac{1}{2} \text{Cov}(G_{\gamma, \text{Mother}}, S_{\gamma, \text{Father}}) + \text{Cov}(\epsilon_G, \epsilon_S)
\]

The linear assortative mating makes any regression you would like to run in the model linear. Exploiting this repeatedly allows us to eliminate all of the non-primitive covariances. I also define the variance of \( W_m \) and \( W_f \) for simplicity of presentation:
\[
\text{Var}(G) = \left( W_G \text{Var}(G) + W_S \text{Cov}(G, S) \right) \left( W_G \text{Var}(G) + \alpha W_S \text{Cov}(G, S) \right) \frac{1}{\sqrt{\text{Var}(W_m) \text{Var}(W_f)}} + 2\text{Var}(\epsilon_G)
\]

(13)

\[
\text{Var}(S) = \left( W_S \text{Var}(S) + W_G \text{Cov}(G, S) \right) \left( W_S \text{Var}(S) + W_G \text{Cov}(G, S) \right) \frac{1}{\sqrt{\text{Var}(W_m) \text{Var}(W_f)}} + 2\text{Var}(\epsilon_S)
\]

(14)

\[
\text{Cov}(G, S) = \left( W_S \text{Var}(S) + W_G \text{Cov}(G, S) \right) \left( W_S \text{Var}(S) + W_G \text{Cov}(G, S) \right) \frac{1}{2\sqrt{\text{Var}(W_m) \text{Var}(W_f)}}
\]

\[
+ \left( W_{S,f} \text{Var}(S) + W_{G,f} \text{Cov}(G, S) \right) \left( W_{S,m} \text{Var}(S) + W_{G,m} \text{Cov}(G, S) \right) \frac{1}{2\sqrt{\text{Var}(W_m) \text{Var}(W_f)}} + \text{Cov}(\epsilon_G, \epsilon_S)
\]

(15)

\[
\text{Var}(W_m) = W_G^2 \text{Var}(G) + W_S^2 \text{Var}(S) + 2W_S W_G \text{Cov}(G, S) + W_G^2 \text{Var}(E) + W_S^2 \text{Var}(E)
\]

(16)

\[
\text{Var}(W_f) = W_G^2 \text{Var}(G) + \alpha W_S^2 \text{Var}(S) + 2\alpha W_S W_G \text{Cov}(G, S) + W_G^2 \text{Var}(E) + \alpha W_S^2 \text{Var}(E)
\]

(17)

With a few simplifying assumptions, I can characterize the model completely.
3.2 Simplified Model

The simplifications we need to make are:

1. $\text{Cov}(E) = 0$, meaning that environmental shocks to the general and systemizing abilities do not covary.

2. $\epsilon_G$ and $\epsilon_S$ are i.i.d., with $\text{Var}(\epsilon_G) = \text{Var}(\epsilon_S) = \text{Var}(\epsilon)$. That rules out coheritability, so if your daughter has more general ability than your son, that does not give you any information about whether she has more systemizing ability.

   I will add a third shortly:

3. $\alpha = 1$ so that the returns to systemizing do not depend on sex.

   Positive assortative mating causes the variance of genetic traits to increase, so the key result I will get below is that increasing the returns to systemizing ($W_S$) will cause the variance of the systemizing distribution to increase ($\text{Var}(S_S)$). Before imposing the third assumption, I want to get a more limited result for $\alpha$. If we are in a genetic steady state, and $\alpha$ increases, what happens to $\text{Var}(S_S)$ in the next generation? I find this by differentiating the steady state condition for $\text{Var}(S_S)$, holding all of the variances and covariances constant. For ease of presentation, I again represent the variance of the male and females wages as $\text{Var}(W_M)$ and $\text{Var}(W_F)[\alpha]$, where the latter needs to be recognized as a function of alpha.
\[ \frac{\partial \text{Var}_{t+1}(S_\gamma)}{\partial \alpha} \bigg|_{\text{Var}(S_\gamma), \text{Var}(G_\gamma), \text{Cov}(G_\gamma, S_\gamma)} = \frac{W_S \text{Var}_t(S_\gamma) (W_S \text{Var}_t(S_\gamma) + W_G \text{Cov}_t(G_\gamma, S_\gamma))}{\sqrt{\text{Var}_t(W_M) \sqrt{\text{Var}_t(W_F)}[\alpha]}} \]

\[ - \alpha W^2 S \text{Var}_t(S_\gamma) (W_S \text{Var}_t(S_\gamma) + W_G \text{Cov}_t(G_\gamma, S_\gamma)) (\alpha W_S \text{Var}_t(S_\gamma) + W_G \text{Cov}_t(G_\gamma, S_\gamma)) \]

\[ \sqrt{\text{Var}_t(W_M) (\text{Var}_t(W_F)[\alpha])^{3/2}} \]

So \[ \frac{\partial \text{Var}_{t+1}(S_\gamma)}{\partial \alpha} \bigg|_{\text{Var}(S_\gamma), \text{Var}(G_\gamma), \text{Cov}(G_\gamma, S_\gamma)} > 0 \] if:

\[ W_S \text{Var}_t(S_\gamma) (W_S \text{Var}_t(S_\gamma) + W_G \text{Cov}_t(G_\gamma, S_\gamma)) \]

\[ \sqrt{\text{Var}_t(W_M) \sqrt{\text{Var}_t(W_F)}[\alpha]} \]

\[ - \alpha W^2 S \text{Var}_t(S_\gamma) (W_S \text{Var}_t(S_\gamma) + W_G \text{Cov}_t(G_\gamma, S_\gamma)) (\alpha W_S \text{Var}_t(S_\gamma) + W_G \text{Cov}_t(G_\gamma, S_\gamma)) \]

\[ \sqrt{\text{Var}_t(W_M) (\text{Var}_t(W_F)[\alpha])^{3/2}} \]

\[ \alpha W_S (\alpha W_S \text{Var}_t(S_\gamma) + W_G \text{Cov}_t(G_\gamma, S_\gamma)) \]

Substituting for \( \text{Var}(W_F) \) and simplifying gives

\[ \alpha < \frac{W^2 G (\text{Var}_t(G_\gamma) + \text{Var}_t(G_E)) + W^2 S \text{Var}_t(S_E) + 2W_G W_S \text{Cov}_t(G_\gamma, S_\gamma)}{W_G W_S \text{Cov}_t(G_\gamma, S_\gamma)} \]

Which is satisfied for all \( \alpha <= 1 \), so that as the female return to systemizing ability moves closer to male return, the variance of the genetic systemizing distribution will increase, at least in the first period.

\[ \frac{\partial \text{Var}_{t+1}(S_\gamma)}{\partial \alpha} \bigg|_{\text{Var}(S_\gamma), \text{Var}(G_\gamma), \text{Cov}(G_\gamma, S_\gamma)} > 0 \]
3.3 Characterizing the Simplified Model

Now we impose all three of the simplifying assumptions, including the third, that $\alpha = 1$. Under these conditions, we can solve for the stationary values:

$$\text{Var}(G) = 2\text{Var}(\epsilon) \left( 1 + \frac{2W_G^2\text{Var}(\epsilon)}{W_G^2(\text{Var}(G_E) - 2\text{Var}(\epsilon)) + W_M^2(\text{Var}(S_e) - 2\text{Var}(\epsilon))} \right)$$

(20)

$$\text{Var}(S) = 2\text{Var}(\epsilon) \left( 1 + \frac{2W_S^2\text{Var}(\epsilon)}{W_G^2(\text{Var}(G_E) - 2\text{Var}(\epsilon)) + W_S^2(\text{Var}(S_E) - 2\text{Var}(\epsilon))} \right)$$

(21)

$$\text{Cov}(G, S) = 2\text{Var}(\epsilon) \left( \frac{2\text{Var}(\epsilon)W_GW_S}{W_G^2(\text{Var}(G_E) - 2\text{Var}(\epsilon)) + W_S^2(\text{Var}(S_E) - 2\text{Var}(\epsilon))} \right)$$

(22)

For comparison, consider random mating, which is defined as non-assortative mating. The stationarity condition for $\text{Var}(G)$ would be:

$$\text{Var}(G) = \text{Var} \left( \frac{1}{2} G_{\gamma, \text{Father}} + \frac{1}{2} G_{\gamma, \text{Mother}} + \epsilon_G \right)$$

(23)

$$= \frac{1}{2} \text{Var}(G) + \text{Var}(\epsilon_G)$$

$$= 2\text{Var}(\epsilon_G)$$

$$= 2\text{Var}(\epsilon)$$

since the covariance between the mother’s and father’s genetic contributions would be zero. So introducing assortative mating makes the variance of a trait
a multiple of the variance under random mating.

To make the results easier to read, let’s define heritability, which is the proportion of the variance of a trait that is due to variation in genes. The heritability\textsuperscript{19} of general ability would be:

\[ h^2(G) = \frac{\text{Var}(G)}{\text{Var}(G)} = \frac{\text{Var}(G)}{\text{Var}(G) + \text{Var}(G_E)} \] (24)

Under random mating, this would be:

\[ h_0^2(G) = \frac{2\text{Var}(\epsilon)}{2\text{Var}(\epsilon) + \text{Var}(G_E)} \] (25)

Using the heritability under random mating for general ability and systemizing, I can rewrite the result as:

\[ \text{Var}(G) = 2\text{Var}(\epsilon) \left( 1 + \frac{W_G^2}{W_G^2 \left( \frac{1}{h_0^2(G)} - 2 \right) + W_S^2 \left( \frac{1}{h_0^2(S)} - 2 \right)} \right) \] (26)

\[ \text{Var}(S) = 2\text{Var}(\epsilon) \left( 1 + \frac{W_S^2}{W_G^2 \left( \frac{1}{h_0^2(G)} - 2 \right) + W_S^2 \left( \frac{1}{h_0^2(S)} - 2 \right)} \right) \] (27)

\[ \text{Cov}(G, S) = 2\text{Var}(\epsilon) \left( \frac{W_G W_S}{W_G^2 \left( \frac{1}{h_0^2(G)} - 2 \right) + W_S^2 \left( \frac{1}{h_0^2(S)} - 2 \right)} \right) \] (28)

The stationary variance under assortative mating is the variance under assortative mating.

\textsuperscript{19}Strictly speaking, this is narrow-sense heritability. From that perspective, I have put the dominance variance into the environmental variance since neither are affected by assortative mating.
random mating, scaled up by a function of the returns to each ability and the
heritability of each trait under random mating.

As the equation shows, \( h_0^2(G) \) and \( h_0^2(S) \) must both be less than \( \frac{1}{2} \), though
\[
\lim_{h_0(G) \to \frac{1}{2}, h_0(S) \to \frac{1}{2}} \text{Var}(G_\gamma) = \infty.
\]
That means that the model can rationalize any observed heritability, and if that observed heritability is greater than \( \frac{1}{2} \), that is due in part to assortative mating with the implied random mating heritability being less than \( \frac{1}{2} \).

3.4 Comparative Statics

I can get comparative statics results for the simplified model:

<table>
<thead>
<tr>
<th></th>
<th>( W_G )</th>
<th>( W_S )</th>
<th>\text{Var}(G_E)</th>
<th>\text{Var}(S_E)</th>
<th>\text{Var}(\epsilon)</th>
</tr>
</thead>
<tbody>
<tr>
<td>\text{Var}(G_\gamma)</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>\text{Var}(S_\gamma)</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>\text{Cov}(G_\gamma, S_\gamma)</td>
<td>+/-</td>
<td>+/-</td>
<td>-</td>
<td>-</td>
<td>+</td>
</tr>
</tbody>
</table>

I can also get firm predictions for the covariance if I assume, for example, that \( \text{Var}(G_E) = \text{Var}(S_E) \), and that \( W_G > W_S \).

<table>
<thead>
<tr>
<th></th>
<th>( W_G )</th>
<th>( W_S )</th>
<th>\text{Var}(G_E)</th>
<th>\text{Var}(S_E)</th>
<th>\text{Var}(\epsilon)</th>
</tr>
</thead>
<tbody>
<tr>
<td>\text{Var}(G_\gamma)</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>\text{Var}(S_\gamma)</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>\text{Cov}(G_\gamma, S_\gamma)</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
</tr>
</tbody>
</table>
Loosely speaking, as the two returns get closer together, the covariance goes up, and as they get farther apart, the covariance falls.

The primary result is that as the returns to an ability increase, the variance of the genetic distribution of that ability will also increase. In footnote 18, I mentioned that you can view $W_G$ and $W_S$ either as labor market returns, in which case it would be complementarities in the household production function that drive assortative mating, or you can view them as marriage market returns to the abilities. The latter approach gives a more direct argument for assortative mating, and allows for more general interpretation, though it makes it harder to match up the $W$ values to anything in real data.

There are also at least two interpretations of $G_E$ and $S_E$. The first is the standard one in quantitative genetics: that there are environmental inputs to traits and that these environmental factors contribute to the true, realized value of the trait. The second is to think of $G_E$ and $S_E$ containing environmental factors that contribute to the true value of abilities, but also as containing noisy variance unrelated to true ability. For example, it is widely agreed that general intelligence comes partially from genes and partially from environmental factors (early child nutrition, schooling, etc). But suppose that there are also some people who appear to everyone as more intelligent than they actually are (and some people who appear less intelligent than they actually are). If mating occurs on perceived intelligence, then this is a kind of search friction in the sense that the matching process gives socially suboptimal matches. Under this interpretation, the comparative statics imply that reduced noise in the either ability results in higher genetic variance for both abilities.
Increased assortative mating causing increased variance was first suggested by Fisher (1918),\textsuperscript{20} and is a result in many subsequent models, including Kremer (1997). Kremer’s basic result is that increased assortative mating will not have too dramatic of an impact on the standard deviation of the distribution. I get a stronger result both by focusing on a trait which is heavily influenced by the parents’ value of the trait (which would give a stronger result in Kremer’s framework), and especially by focusing on the tails of the distribution which magnifies the effect of any change in the variance.

3.5 Extreme Outcomes and Normal Tails

The central reason that increased assortative mating could explain even a large increase in the prevalence of ASDs is that, under the systemizing view, ASDs are the extreme right tail of the systemizing distribution. To show why focusing on tails magnifies the effect of increased variance, I will find the elasticity of the tail area of a normal distribution with respect to the standard deviation, and I will to do this for different definitions of the tail. That is, for a threshold $T$ and a standard deviation $\sigma$, I will show how much the area past $T$ increases as I increase $\sigma$.

\textsuperscript{20}As a historical note, Fisher’s paper is the first to use the word “variance” to denote the square of the standard deviation.

\textsuperscript{21}Fernandez and Rogerson (2001) find a stronger effect than Kremer, though their approach is very different from both Kremer’s and mine.
That is

\[
\eta_\sigma(T, \sigma) := \frac{\sigma}{\text{Area}(T, \sigma)} \frac{\partial}{\partial \sigma} \left[ \text{Area}(T, \sigma) \right] \tag{29}
\]

\[
= \frac{\sigma}{1 - \frac{1}{\sqrt{2\pi\sigma}} \int_0^T e^{-\frac{x^2}{2\sigma^2}} \, dx} \frac{\partial}{\partial \sigma} \left[ 1 - \frac{1}{\sqrt{2\pi\sigma}} \int_0^T e^{-\frac{x^2}{2\sigma^2}} \, dx \right]
\]

\[
= \sqrt{\frac{2T}{\pi\sigma}} \frac{\frac{e^{-\frac{T^2}{2\sigma^2}}}{\sqrt{\pi\sigma^3}}}{1 - \frac{2}{\sqrt{\pi}} \int_0^T e^{-x^2} \, dx}
\]

\[
= \eta_\sigma \left( \frac{T}{\sigma} \right)
\]

This does not have closed form due to the integral, but notice that it is homogenous of degree zero in $T$ and $\sigma$. This allows us to treat the elasticity as a function of the threshold $T$ stated in units of $\sigma$.

To give some intuition, consider first evaluating this at $\frac{T}{\sigma} = 0$, that is, how does the area of the top half of the distribution change when you increase the standard deviation? Looking at the formula, you see that $\eta_\sigma(0) = 0$, which is just to say that if I increase the standard deviation, there is still exactly half the mass of a normal distribution to the right of the mean. But now consider $\eta_\sigma(1)$, which you can see from the formula is greater than 0. Imagine the area past one standard deviation to right of the mean (this is the familiar .159). Now if you increase the standard deviation from $\sigma$ to $\tilde{\sigma}$, how much area is to the right of $\mu + \sigma$? Clearly, it must be more than .159, because that is the area which is now to the right of $\mu + \tilde{\sigma}$ which is greater than $\mu + \sigma$. That shows that $\eta_\sigma(1) > 0$.  

23
Below are numerical values for $\eta_\sigma$:

![Graph showing elasticity vs. threshold in units of $\sigma$.]

As an example, suppose that those well-qualified for systemizing jobs like being an engineer (or economist) are those who are more than one-standard deviation above the mean in the systemizing distribution, and $\eta_\sigma(1) = 1.53$. If full-blown autism had a prevalence of around .001 a generation ago, and those with autism represent the extreme right tail of the systemizing distribution, then those with full-blown autism were 3.09 standard deviations away from the mean, and $\eta_\sigma(3.09) = 10.4$. So $\frac{\eta_\sigma(3.09)}{\eta_\sigma(1)} = \frac{10.4}{1.53} = 6.82$, which means that increases in the variance of the systemizing distribution will increase the autistic population nearly seven times as much (in proportional terms) as it will increase the population of those qualified for systemizing jobs. Therefore, a shift in assortative that doubled the size of the autistic population would cause only a $\frac{100\%}{6.82} = 14.7\%$ increase in the population of those qualified for systemizing jobs.

As another example, take the height distribution. The CDC runs a recurring study called the National Health and Nutrition Examination Survey.
(NHANES) which collects a variety of information, including height in a na-
tionally representative sample, and versions of the study have existed since
1959.\footnote{The earliest study was called the National Health and Examination Survey (NHES),
though it is comparable for the variables I am interested in.} This theory says that if the height distribution is approximately nor-
mal, then knowing the percent change in the standard deviation should be
sufficient to figure out the percent growth beyond any threshold. Let’s take
as an example those above six feet tall, and look at those between age 20 and
age 55.

In 1959, 235 of 4926 people sampled were taller than 72 inches (six feet),
which is 4.7\% of the population. In 2009, 310 of 3712 people sampled were
taller than 72 inches, which is 8.35\% of the population, 75\% higher than in
1959. However, part of this is due to a general increase in height: between
1959 and 2009, the mean height of those between age 20 and age 55 increased
by .53 inches. To account for the increase in the mean, I add .53 inches to the
2009 thresholds. In 2009, 241 of 3712 people sampled were taller than 72.53
inches, which is 6.5\% of the population, and 36\% higher than 1959.

To compute our prediction for comparison with the true value, we need to
know how many standard deviations 72 inches was from the mean in 1959 (1.67
standard deviations), and how much the standard deviation grew between
1959 and 2009 (10.35\%). The formula gives $n_\sigma(1.67) = 3.47$, so our predicted
increase would be $3.47 \times 10.35\% = 35.9\%$ which is very close to the observed
36\%. This is a particularly lucky example, but the chart below shows the
actual growth at one-inch intervals compared to the blue prediction line (the
72 inch example is the third from the right) and it performs well across the
distribution.

To be clear, this is not a deep test of the complete theory I have presented: all that is needed is for the height distribution to be normal in both 1959 and 2009. However, it is reassuring that the tails do grow at the high rates predicted, and the increase in the standard deviation is consistent with increased assortative mating. The percent of those below 58 inches (4’10”) grew 8.8% between 1959 and 2009, and that is without correcting for the increased mean (it grew 97% once I account for the mean). Growth at the bottom of the distribution is hard to explain through factors like nutrition that we typically discuss when thinking about long-term changes in the height distribution, but it is very easy to explain in an assortative mating framework.

3.6 Potential Effect Size

Having seen in section 3.5 that relatively small changes in mating could have large effects on the extremes of distributions, I want to present a few exam-
ples of how specific changes in mating would affect the prevalence of autism. These are reduced form estimates that just use the correlation in systemizing among spouses. Suppose that we were in a genetic steady state with a spousal systemizing correlation of .3, and at generation 1, the correlation shifted to .4. Below is how autism prevalence would evolve over the ensuing generations, starting from the typical estimate of 1 in 1000 prevalence.

This relatively modest change in mating causes a doubling in autism rates, though it does so over several generations. But another way to look at that is that effects from changes in mating will continue to be felt generations later, so that shifts in mating which occurred several generations ago could be contributing to the current rise in autism.
Here is the chart if the correlation moved from .2 to .4

This more dramatic change would cause a 50% increase in prevalence in a single generation, and an eventual tripling.

4 Assortative Mating

If autism is the extreme right tail of the systemizing distribution, section 3 gives us a possible explanation for rising ASD prevalence: increased mating on systemizing has increased the variance of the systemizing distribution, and dramatically increased the tail area in relative terms. The role of assortative mating in general is suggested by Baron-Cohen,\textsuperscript{23} and has also appeared in the popular press as a possible explanation for the high rates of Asperger’s syndrome among children in Silicon Valley.\textsuperscript{24} However, an increase in marriages

\textsuperscript{23}e.g. Baron-Cohen, (2006)
\textsuperscript{24}Silberman (2001)
of systemizers has not been empirically documented, and the role of the tail in magnifying the impact of assortative mating has not been recognized. There are many more women in systemizing occupations than there used to be, and we see more marriages where husband and wife are both in systemizing occupations. However, it is possible that the schoolteacher wife of a male engineer from a few generations ago would have been an engineer herself if she’d had that opportunity. If that were the case, it could be that assortative mating on observables rose, but assortative mating on genes held completely steady. I will first document the increase in sorting on observables, and then present evidence that mating has meaningfully shifted.

4.1 Matching on Degrees

I’ll approach this by using the 2010 American Community Survey. The survey asked respondents for the field of their college degree beginning in 2009, and this gives us a way to look at systemizing using the cross-section. The census classifies fields of study into broad and specific categories. One of the broad categories is science and engineering degrees, which I’ll use as a measure of systemizing ("Census Science"). The census’s definition includes degrees in psychology and in the social sciences (both of which are specific categories), so I’ll also use a narrow measure which excludes those two ("Hard Science")\(^{25}\).

The data for men and women look very different over the period. Here is male degree type by age (with age running backwards so that more recent cohorts are to the right).

\(^{25}\)The specific categories that remain are computers and math, biological sciences, physical science, engineering, and multidisciplinary studies.
There are two things to observe in the graph. First, the graph supports the idea that field of degree is a meaningful measure. The spike in degree attainment around age 60 is driven by the Vietnam war and the associated draft.\textsuperscript{26} The spike is much more muted when we look at census science degrees, and is basically absent when we look at hard science degrees. If field choice were purely random, we would see a proportional spike in each line, so the absence of a spike suggests that hard science degrees require certain abilities that are not common among those on the margin for college. Second, the portion of men with hard science degrees has remained nearly constant, at

\textsuperscript{26}Card and Lemieux (2001)
just under 10%.

Women experienced a steep rise in the change of graduating college, and an even steeper rise in the chance of getting a science degree. For example, 1.1% of women above age 70 got a hard science degree, while 6% of women between ages 30 and 35 got a hard science degree.\textsuperscript{27}

Sorting on observables has risen steadily. Below is the percentage of married couples who both have a hard science degree, by the wife’s age.

\textsuperscript{27}Because human capital increases longevity, it is possible that the estimate for women above 70 is biased upward, which would cause the growth rate to be underestimated.
4.2 Matching Across States

Clearly some of the rise in sorting on observables is driven by the increasing number of women with science degrees. The important question is whether those women are marrying in a different way than women who had the same (latent) ability before them.

We can see evidence of an actual change in mating by looking at the tendency of people to marry someone born in their birth state. For woman married since 1960, I have listed the probability of a birth state match by the type of her degree, for couples who were both born in the United States.
The overall probability of marrying someone born in your state has been falling over time.

The way that the degree-specific probabilities evolve over time will give an indication of whether women’s marriage behavior depends on observed ability or on latent ability. If latent ability is what matters, then the probability for each observable category is a weighted average of the true, latent probabilities.
Then, as women’s latent abilities become more observable over time, the observed probabilities will move apart. By contrast, if observed ability is what matters, then the probabilities should move roughly in parallel, because as women’s latent ability becomes observable, their behavior changes to match the behavior of women who already had that observed ability.

To make this clearer, imagine just having college graduates and non-graduates. As more women graduate college, what happens to the probability of a birth state match for non-graduates? If what matters is the observable outcome (that is, not graduating college), then nothing will happen. The women who now graduate college begin acting like college graduates when they become college graduates.

But if latent ability is what matters, the story is different. Some women who did not graduate college had the latent ability to do so, so they were already marrying like college graduates. As more women attend college, the probability of a birth state match for non-graduates will rise, as the observed category becomes a more accurate representation of the latent category.

Below I calculate how much I would expect the lines to diverge, if latent ability is what mattered. I am focusing on just three categories (no degree, non-science degrees, and census science degrees) because the census science degrees and hard science degrees track each other very closely. I chart all, including hard science degrees, in appendix A.

If latent ability is what matters, then the probabilities we initially see are weighted averages of the true probabilities. For now, I am ignoring the time trend in the overall probabilities, though I discuss this below.
<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Share of women with no degree in 1960-62 ( (\alpha_{\text{No Degree}}) )</td>
<td>.814</td>
</tr>
<tr>
<td>Share of women with no degree in 2005-07 ( (\alpha'_{\text{No Degree}}) )</td>
<td>.616</td>
</tr>
<tr>
<td>Share of women with a non-science degree in 1960-62 ( (\alpha_{\text{Non-Sci}}) )</td>
<td>.159</td>
</tr>
<tr>
<td>Share of women with a non-science degree in 2005-07 ( (\alpha'_{\text{Non-Sci}}) )</td>
<td>.274</td>
</tr>
<tr>
<td>Share of women with a science degree in 1960-62 ( (\alpha_{\text{Science}}) )</td>
<td>.026</td>
</tr>
<tr>
<td>Share of women with a science degree in 2005-07 ( (\alpha'_{\text{Science}}) )</td>
<td>.11</td>
</tr>
</tbody>
</table>

We will also need the initial, observed birth state match probabilities for each type.

<table>
<thead>
<tr>
<th>Degree type</th>
<th>Initial birth state match probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>No degree ( (\beta_{\text{No Degree}}) )</td>
<td>.667</td>
</tr>
<tr>
<td>Non-science degree ( (\beta_{\text{Non-Sci}}) )</td>
<td>.561</td>
</tr>
<tr>
<td>Science degree ( (\beta_{\text{Science}}) )</td>
<td>.514</td>
</tr>
</tbody>
</table>

The goal is to recover \( \gamma_{\text{No Degree}}, \gamma_{\text{Non-Sci}}, \) and \( \gamma_{\text{Science}}, \) which are the birth state match probabilities for women who have the latent ability indicated by the subscripts.

To write out equations for the other \( \beta \)'s, we will need to know the fraction of women who eventually got science degrees who counterfactually would have initially gotten a non-science degree (as opposed to no degree). I will let this quantity vary, but I will call it \( \eta \). So:
\[ \beta_{\text{Science}} = \gamma_{\text{Science}} \]

\[ \beta_{\text{Non-Sci}} = \frac{(\alpha_{\text{Non-Sci}} - \eta(\alpha'_{\text{Science}} - \alpha_{\text{Science}})) \gamma_{\text{Non-Sci}} + \eta (\alpha'_{\text{Science}} - \alpha_{\text{Science}}) \gamma_{\text{Science}}}{\alpha_{\text{Non-Sci}}} \]

\[ \beta_{\text{No Degree}} = \frac{\alpha'_{\text{No Degree}} \gamma_{\text{No Degree}} + (1 - \eta)(\alpha'_{\text{Science}} - \alpha_{\text{Science}}) \gamma_{\text{Science}}}{\alpha_{\text{No Degree}}} + \frac{(\alpha'_{\text{Non-Sci}} - \alpha_{\text{Non-Sci}} + \eta(\alpha'_{\text{Science}} - \alpha_{\text{Science}})) \gamma_{\text{Non-Sci}}}{\alpha_{\text{No Degree}}} \]

These equations can easily solved for the \( \gamma \)'s. Because we have let \( \eta \) vary, we will have \( \gamma_{\text{Science}}(\eta) \), \( \gamma_{\text{Non-Sci}}(\eta) \), and \( \gamma_{\text{No Degree}}(\eta) \). Since the probability for science degrees does not change, the percent growth in the distance between the science and non-science probabilities would be:

\[ \frac{\gamma_{\text{Non-Sci}}(\eta) - \beta_{\text{Non-Sci}}}{\beta_{\text{Non-Sci}} - \beta_{\text{Science}}} \]

The percent growth in the distance between the no degree and non-science probabilities would be:

\[ \frac{\gamma_{\text{No Degree}}(\eta) - \beta_{\text{No Degree}} - (\gamma_{\text{Non-Sci}}(\eta) - \beta_{\text{Non-Sci}})}{\beta_{\text{No Degree}} - \beta_{\text{Non-Sci}}} \]
As you can see from the projected charts, if latent ability mattered, the lines should have moved noticeably apart. At very low values of $\eta$, we would
have seen significant growth between the no degree and non-science lines. At any other level of $\eta$, we would see growth between the non-science and science lines.

Below are the actual probabilities over time.

The probabilities track each other closely, instead of diverging as they would if latent ability determined behavior. This suggests that women’s abilities becoming more observable actually changed marriage behavior.

It is possible that the compositional changes would have caused the lines to diverge, but the true probabilities shifted and brought the lines back together. However, this would require the less educated to become relatively more mobile, and the more educated to become relatively less mobile.
5 Welfare and Policy

One response to the idea that increased assortative mating on systemizing has increased the autistic population would be to discourage systemizers from marrying. However, I believe that this would be misguided. Despite the high costs of severe autism, increased assortative mating will also provide social benefits. Below, I will do a rough calculation to show that the net benefits from increased assortative mating on systemizing are likely positive.

To do this, I am going to take a very narrow look at the benefits and a very broad look at the costs, and I will find that the benefits outweigh the costs. For the benefits, I will just be looking at labor market earnings. Increased assortative mating will increase the variance of ability, so if the return to systemizing is constant across the distribution, all the gains will cancel the losses. However, the returns to ability are far from constant, which I will illustrate with the NLSY79. Almost all respondents took the Armed Services Vocational Aptitude Battery (ASVAB) around 1979 (when they were in their late teens). I will use their scores on the three math sections of the ASVAB as a measure of systemizing, and earnings in their 2010 jobs (conditional on working) to get a quick estimate of the marginal return to systemizing\(^{28}\). Below are the smoothed means:

\(^{28}\)Including a measure of general ability makes the cost-benefit result stronger, so I have omitted it in the spirit of making the most conservative assumptions.
I assume a quadratic relationship, and run separate regressions for those above and below the mean. The math measure is demeaned and divided by its standard deviation, so that the integral below can be indexed by the standard deviation. The regressions themselves are in appendix B, and are of the form:

\[
\text{Income}_{2010} = \begin{cases} 
\alpha_{\text{Below}} + \beta_{\text{Below Math}} + \gamma_{\text{Below Math}}^2 + \varepsilon & \text{if Math} \leq \text{Mean(Math)} \\
\alpha_{\text{Above}} + \beta_{\text{Above Math}} + \gamma_{\text{Above Math}}^2 + \varepsilon & \text{if Math} > \text{Mean(Math)}
\end{cases}
\]

I will also need to estimate costs along the autistic spectrum. Ganz (2007) estimated the lifetime social costs of a marginal case of autism, including med-
atical and non-medical care, and lost productivity, and arrived at $3.2 million.\textsuperscript{29} To be generous, let’s call it $5 million.\textsuperscript{30} Ganz’s estimate is more tilted towards severe cases, so we also need an estimate for people with Asperger’s syndrome (milder autism). Given how many high achievers have Asperger’s, I believe that there is a strong argument that the costs of having Asperger’s are below the earnings potential of someone with Asperger’s. However, to be conservative, I will assume that the cost of having Asperger’s completely cancels any labor market earnings.

Now I can make a rough welfare calculation. To get lifetime earnings, I multiply the earnings integrals by 30 years. I use the standard estimates of autism being a 3 standard deviation trait, and Asperger’s being a 2.5 standard deviation trait. Below is the net benefit (or cost) of increasing the standard deviation of systemizing by 10%.

\textsuperscript{29}Ganz (2007)
\textsuperscript{30}I am also implicitly raising the estimate again in my calculation below. Ganz’s estimate already includes lost productivity, and I assume zero productivity and then take Ganz’s estimate as a cost
Net benefit = $30 \int_{-\infty}^{0} \left( \hat{\beta}_{\text{Below}} x + \hat{\gamma}_{\text{Below}} x^2 \right) \left( \frac{1}{1.1 \sqrt{2\pi}} e^{\frac{x^2}{2(1.1)^2}} - \frac{1}{1.1 \sqrt{2\pi}} e^{\frac{x^2}{2(1)^2}} \right) dx$

Lower lifetime earnings on the bottom half of the distribution

$+ 30 \int_{0}^{2.5} \left( \hat{\beta}_{\text{Above}} x + \hat{\gamma}_{\text{Above}} x^2 \right) \left( \frac{1}{1.1 \sqrt{2\pi}} e^{\frac{x^2}{2(1.1)^2}} - \frac{1}{1.1 \sqrt{2\pi}} e^{\frac{x^2}{2(1)^2}} \right) dx$

Higher lifetime earnings on the top half of the distribution

$+ \int_{2.5}^{3} 0 \left( \frac{1}{1.1 \sqrt{2\pi}} e^{\frac{x^2}{2(1.1)^2}} - \frac{1}{1.1 \sqrt{2\pi}} e^{\frac{x^2}{2(1)^2}} \right) dx$

Asperger's is net neutral

$+ \int_{3}^{\infty} -5,000,000 \left( \frac{1}{1.1 \sqrt{2\pi}} e^{\frac{x^2}{2(1.1)^2}} - \frac{1}{1.1 \sqrt{2\pi}} e^{\frac{x^2}{2(1)^2}} \right) dx$

The lifetime cost of each new case of autism

$= -$21,869

Lower lifetime earnings on the bottom half of the distribution

$+ $42,217

Higher lifetime earnings on the top half of the distribution

$+ $0

Asperger's is net neutral

$+ -$12,896

The lifetime cost of each new case of autism

$= $11,132

So the increased expected wages by themselves outweigh the expected costs
of additional cases of autism. While the increased risk of autism is probably not priced into mating decisions, this suggests that increased assortative mating on systemizing is probably a good thing on balance.

But while direct intervention does not seem like a good idea, there are still policy implications to this view of autism. If assortative mating has significantly contributed to the rise in autism, autism is likely to keep rising for several more generations, because it takes multiple generations to reach a new genetic equilibrium. Investments in better autism care are probably more cost-effective than they appear at the current prevalence.

6 Conclusion

This paper explored assortative mating’s impact on the extremes of distributions through an application to autism. I presented a general model of genes and mating, where mating is determined by returns to abilities that are generated by genes and environment. I find that increased returns to a trait cause more assortative mating on that trait, and that increasing the returns to a trait for women alone will also increase assortative mating.

To make the connection to autism, I provided the first large sample, general population evidence for a connection between systemizing and autism. That link does not seem to be caused by either socio-economics status or the detection methodology. I have shown that relatively modest shifts in assortative mating can produce very large impacts, and I calculated that if the spousal correlation of systemizing moved from .3 to .4, autism prevalence would even-
tually double. A shift in mating that caused autism prevalence to double would only increase the share of the population past one standard deviation in the systemizing distribution by under 15%, so autism could have increased significantly without dramatically changing the overall ability distribution.

My model suggests that developments like computers, which have likely increased the returns to systemizing, would lead to more assortative mating on systemizing. The same is true for the shifts that have caused women to enter systemizing occupations at much higher rates. I find evidence that mating on observable systemizing has increased, and that the increase reflects a true shift in who marries whom, and is not entirely driven by women’s latent systemizing ability becoming observable.

Taken together, this paper shows that assortative mating could have played a significant role in the rise of autism rates, and more generally, that even in cases where assortative mating does not dramatically change the population distribution, small populations at the extremes may be dramatically affected.
References


Newschaffer, Craig J., Lisa A. Croen, Julie Daniels, Ellen Giarelli, Judith K. Grether, Susan E. Levy, David S. Mandell, Lisa A.


Appendix A: Birth State Match Chart with Hard Science Degrees

Here is the birth state match probability chart which includes hard science degrees. Hard science degrees track census science degrees very closely, but have larger confidence intervals, and would present one more category and two more parameters for the calculation in section 4.2. Since the probabilities are so similar, the additional structure would not significantly alter the analysis.
Appendix B: Regressions for Section 5

Below are the regressions for the rough welfare calculation in section 5. As discussed in the text, the data come from the NLSY79. I combine the Arithmetic Reasoning (AR), Mathematics Knowledge (MK), and Numerical Operations (NO) scores to get a combined mathematics score. For everything that follows, including the regressions, I use the 2010 cross-sectional weights.

I demean the score and divide it by its standard deviation so that the regression coefficients will give returns per standard deviation, since the integral is indexed by standard deviations.

For earnings, I take labor market earnings from 2010 jobs (conditional on working).
<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Income if Math&gt;mean</td>
<td>15,036**</td>
<td>-19,234</td>
</tr>
<tr>
<td></td>
<td>(6,155)</td>
<td>(12,493)</td>
</tr>
<tr>
<td>Income if Math&lt;mean</td>
<td>-1,230</td>
<td>32,207***</td>
</tr>
<tr>
<td></td>
<td>(3,184)</td>
<td>(6,951)</td>
</tr>
<tr>
<td>ASVAB Math</td>
<td>48,771***</td>
<td>58,133***</td>
</tr>
<tr>
<td></td>
<td>(2,496)</td>
<td>(4,653)</td>
</tr>
<tr>
<td>ASVABMath$^2$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>1,432</td>
<td>1,781</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.034</td>
<td>0.080</td>
</tr>
</tbody>
</table>

ASVAB Math scores are in standard deviation units

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1