Childhood Autism and Assortative Mating

Hays Golden*

University of Chicago

April 22, 2013

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 Preventions (CDC) Metropolitan Atlanta Developmental Disabilities Surveillance Program (MADDSP), 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.[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.[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

[1] Thiessen and Gregg (1980)

[2] I discuss this more formally in section 2.6.

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 2.6 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 2.6, 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 presenting a general model of endogenous assortative mating on traits that are at least partially passed from parents to children. The model, which has some similaritites to Becker and Tomes (1979) and Kremer (1997), shows that the assortative mating has a particularly strong impact on how often we see extremes levels of traits. Next, I extend the evidence for a connection between ASDs and systemizing in a large sample, general population dataset in section 3. In section 4, I present evidence that ASD rates have actually risen. I then provide some evidence that mating has changed, in section 5. Finally, in section 6, I do a rough calculation which suggests that increased assortative mating on systemizing has probably been socially beneficial, despite the increase in autism rates.

2 Assortative Mating and Genetic Distributions

To better understand the impact of endogenous assortative mating across generations, I construct a formal model that shows how changing returns to abilities ultimately affect the distributions of those abilities. I build the model by adding a marriage market to 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

does not contribute to individual dynamics as it will in my model.[5] 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 will have a "wage," W, which comes from their levels of differents traits. To construct the marriage market, I assume that the marital surplus function is supermodular in W, which will generate strict positive assortative mating on W.[6] There are several channels that could lead to stronger positive assorting mating on a trait like systemizing, including a greater taste among systemizers for systemizing in mates, or decreased search frictions. However, I believe that taking the single index approach and modeling mating on W has several advantages.

First, the effect that I'm modeling is extremely general, and will exist even in situations where there isn't multidimensional attraction (i.e. where your taste for a trait doesn't depend on your level of that trait). If everyone suddenly decided that a certain trait (say, green eyes) is very desirable, this model implies that we'll see more green-eyed-and-green-eyed couples simply

[5]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_γ) as it would be in that interpretation of Becker and Tomes.

[6]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 as the marriage market return to each agent's abilities. 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.

because people with green eyes will be more able to "afford" other green-eyed people on the marriage market. If any trait becomes more attractive, the logic of mating on status will create more mating on that trait, even if there's no multi-dimensional attraction.

Second, and relatedly, the index approach allows one trait or another to become more important in choosing a partner without preferences changing. If we view W as labor market returns, it also gives us model primatives that are more easily observable than preference parameters.

The effects predicted by this model may be augmented by multi-dimensional forces, or improved search, but this model will provide a very general baseline.

2.1 Model

In the model, agents have two traits, systemizing ability (S) and general ability (G)[7]. Each trait comes partially from genes and partially from environmental factors.

$$S = \underbrace{S_\gamma}_{Genetic} + \underbrace{S_E}_{Environmental} \tag{1}$$

$$G = \underbrace{G_\gamma}_{Genetic} + \underbrace{G_E}_{Environmental} \tag{2}$$

Genetic contributions are normally distributed.

[7]Nothing about the mechanics of the model is specific to these two traits, though the model is probably most natural with one "numeraire" trait like general ability.

$$\begin{pmatrix} G_\gamma \\ S_\gamma \end{pmatrix} \sim N \left[\begin{pmatrix} \mu_G \\ \mu_S \end{pmatrix}, \begin{pmatrix} \sigma^2_{G,\gamma} & Cov_\gamma \\ Cov_\gamma & \sigma^2_{S,\gamma} \end{pmatrix} \right] \tag{3}$$

Environmental contributions are also normally distributed.

$$\begin{pmatrix} G_E \\ S_E \end{pmatrix} \sim N \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2_{G,E} & Cov_E \\ Cov_E & \sigma^2_{S,E} \end{pmatrix} \right] \tag{4}$$

Agents have a wage which comes from their level of each ability, and the returns to systemizing ability may differ by sex.

$$\begin{aligned} W_m =& G \cdot w_G + S \cdot w_S \\ =& (G_\gamma + G_E) w_G + (S_\gamma + S_E) w_S \end{aligned} \tag{5}$$

$$\begin{aligned} W_f =& (G_\gamma + G_E) w_G + (S_\gamma + S_E) \alpha w_S \\ 0 \leq& \alpha \leq 1 \end{aligned} \tag{6}$$

Each couple has one son and one daughter, whose genes are averages of their parents, plus a shock.

$$G_\gamma = \frac{1}{2}G_{\gamma,Father} + \frac{1}{2}G_{\gamma,Mother} + \epsilon_G \tag{7}$$

$$S_\gamma = \frac{1}{2}S_{\gamma,Father} + \frac{1}{2}S_{\gamma,Mother} + \epsilon_S \tag{8}$$

Where

$$\begin{pmatrix} \epsilon_G \\ \epsilon_S \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \text{Var}(\epsilon_G) & \text{Cov}(\epsilon_G, \epsilon_s) \\ \text{Cov}(\epsilon_G, \epsilon_S) & \text{Var}(\epsilon_S) \end{pmatrix}\right] \tag{9}$$

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.

Each couple has a son and a daughter, and each child gets an independent joint draw of ϵ_G and ϵ_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.

2.2 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:

$$\begin{aligned}\text{Var}(G_\gamma) =& \text{Var}\left(\frac{1}{2}G_{\gamma,Father} + \frac{1}{2}G_{\gamma,Mother} + \epsilon_G\right) \\ =& \frac{1}{2}\text{Var}(G_\gamma) + \frac{1}{2}\text{Cov}(G_{\gamma,Father}, G_{\gamma,Mother}) + \text{Var}(\epsilon_G)\end{aligned} \tag{10}$$

$$\begin{aligned}\text{Var}(S_\gamma) =& \text{Var}\left(\frac{1}{2}S_{\gamma,Father} + \frac{1}{2}S_{\gamma,Mother} + \epsilon_S\right) \\ =& \frac{1}{2}\text{Var}(S_\gamma) + \frac{1}{2}\text{Cov}(S_{\gamma,Father}, S_{\gamma,Mother}) + \text{Var}(\epsilon_S)\end{aligned} \tag{11}$$

$$\begin{aligned}\text{Cov}(G_\gamma, S_\gamma) =& \text{Cov}(\frac{1}{2}G_{\gamma,Father} + \frac{1}{2}G_{\gamma,Mother} + \epsilon_G, \frac{1}{2}S_{\gamma,Father} + \frac{1}{2}S_{\gamma,Mother} + \epsilon_S) \\ =& \frac{1}{2}\text{Cov}(G_\gamma, S_\gamma) + \frac{1}{4}\text{Cov}(G_{\gamma,Father}, S_{\gamma,Mother}) \\ &+ \frac{1}{4}\text{Cov}(G_{\gamma,Mother}, S_{\gamma,Father}) + \text{Cov}(\epsilon_G, \epsilon_S) \\ =& \frac{1}{2}\text{Cov}(G_{\gamma,Father}, S_{\gamma,Mother}) + \frac{1}{2}\text{Cov}(G_{\gamma,Mother}, S_{\gamma,Father}) + \text{Cov}(\epsilon_g, \epsilon_S)\end{aligned} \tag{12}$$

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:

$$\mathrm{Var}(G_\gamma) = \frac{(w_G\mathrm{Var}(G_\gamma) + w_S\mathrm{Cov}(G_\gamma, S_\gamma))\,(w_G\mathrm{Var}(G_\gamma) + \alpha w_S\mathrm{Cov}(G_\gamma, S_\gamma))}{\sqrt{\mathrm{Var}(W_m)\mathrm{Var}(W_f)}} + 2\mathrm{Var}(\epsilon_G) \quad (13)$$

$$\mathrm{Var}(S_\gamma) = \frac{(w_S\mathrm{Var}(S_\gamma) + w_G\mathrm{Cov}(G_\gamma, S_\gamma))\,(\alpha w_S\mathrm{Var}(S_\gamma) + w_G\mathrm{Cov}(G_\gamma, S_\gamma))}{\sqrt{\mathrm{Var}(W_m)\mathrm{Var}(W_f)}} + 2\mathrm{Var}(\epsilon_S) \quad (14)$$

$$\begin{aligned}\mathrm{Cov}(G_\gamma, S_\gamma) =& \frac{(w_S\mathrm{Var}(S_\gamma) + w_G\mathrm{Cov}(G_\gamma, S_\gamma)(w_G\mathrm{Var}(G_\gamma) + \alpha w_S\mathrm{Cov}(G_\gamma, S_\gamma))}{2\sqrt{\mathrm{Var}(W_m)\mathrm{Var}(W_f)}} \\ &+ \frac{(\alpha w_S\mathrm{Var}(S_\gamma) + w_G\mathrm{Cov}(G_\gamma, S_\gamma))(w_S\mathrm{Var}(S_\gamma) + w_G\mathrm{Cov}(G_\gamma, S_\gamma))}{2\sqrt{\mathrm{Var}(W_m)\mathrm{Var}(W_f)}} + \mathrm{Cov}(\epsilon_G, \epsilon_S)\end{aligned} \quad (15)$$

$$\mathrm{Var}(W_m) = w_G^2\mathrm{Var}(G_\gamma) + w_S^2\mathrm{Var}(S_\gamma) + 2w_Sw_G\mathrm{Cov}(G_\gamma, S_\gamma) + w_G^2\mathrm{Var}(G_E) + w_S^2\mathrm{Var}(S_E) \quad (16)$$

$$\mathrm{Var}(W_f) = w_G^2\mathrm{Var}(G_\gamma) + \alpha w_S^2\mathrm{Var}(S_\gamma) + 2\alpha w_Sw_G\mathrm{Cov}(G_\gamma, S_\gamma) + w_G^2\mathrm{Var}(G_E) + \alpha w_S^2\mathrm{Var}(S_E) \quad (17)$$

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

2.3 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. ϵ_G and ϵ_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 ($\text{Var}(S_\gamma)$). Before imposing the third assumption, I want to get a more limited result for α. If we are in a genetic steady state, and α increases, what happens to $\text{Var}(S_\gamma)$ in the next generation? I find this by differentiating the steady state condition for $\text{Var}(S_\gamma)$, 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.

$$\frac{\partial \mathrm{Var}_{t+1}(S_\gamma)}{\partial \alpha}\Big|_{\mathrm{Var}_t(S_\gamma),\mathrm{Var}_t(G_\gamma),\mathrm{Cov}_t(G_\gamma,S_\gamma)}$$
$$= \frac{w_S \mathrm{Var}_t(S_\gamma)\left(w_S \mathrm{Var}_t(S_\gamma) + w_G \mathrm{Cov}_t(G_\gamma, S_\gamma)\right)}{\sqrt{\mathrm{Var}_t(W_m)}\sqrt{\mathrm{Var}_t(W_f)[\alpha]}} \tag{18}$$
$$- \frac{\alpha w_S^2 \mathrm{Var}_t(S_\gamma)\left(w_S \mathrm{Var}_t(S_\gamma) + w_G \mathrm{Cov}_t(G_\gamma, S_\gamma)\right)\left(\alpha w_S \mathrm{Var}_t(S_\gamma) + w_G \mathrm{Cov}_t(G_\gamma, S_\gamma)\right)}{\sqrt{\mathrm{Var}_t(W_m)}\left(\mathrm{Var}_t(W_f)[\alpha]\right)^{3/2}}$$

So $\frac{\partial \mathrm{Var}_{t+1}(S_\gamma)}{\partial \alpha}\Big|_{\mathrm{Var}_t(S_\gamma),\mathrm{Var}_t(G_\gamma),\mathrm{Cov}_t(G_\gamma,S_\gamma)} > 0$ if:

$$\frac{w_S \mathrm{Var}_t(S_\gamma)\left(w_S \mathrm{Var}_t(S_\gamma) + w_G \mathrm{Cov}_t(G_\gamma, S_\gamma)\right)}{\sqrt{\mathrm{Var}_t(W_m)}\sqrt{\mathrm{Var}_t(W_f)[\alpha]}}$$
$$- \frac{\alpha w_S^2 \mathrm{Var}_t(S_\gamma)\left(w_S \mathrm{Var}_t(S_\gamma) + w_G \mathrm{Cov}_t(G_\gamma, S_\gamma)\right)\left(\alpha w_S \mathrm{Var}_t(S_\gamma) + w_G \mathrm{Cov}_t(G_\gamma, S_\gamma)\right)}{\sqrt{\mathrm{Var}_t(W_m)}\left(\mathrm{Var}_t(W_f)[\alpha]\right)^{3/2}} > 0$$
$$\alpha w_S(\alpha w_S \mathrm{Var}_t(S_\gamma) + w_G \mathrm{Cov}_t(G_\gamma, S_\gamma) > \sqrt{\mathrm{Var}_t(W_f)[\alpha]}$$

Substituting for $\mathrm{Var}(W_f)$ and simplifying gives

$$\alpha < \frac{w_G^2(\mathrm{Var}_t(G_\gamma) + \mathrm{Var}_t(G_E)) + w_S^2 \mathrm{Var}_t(S_E) + 2 w_G w_S \mathrm{Cov}_t(G_\gamma, S_\gamma)}{w_G w_S \mathrm{Cov}_t(G_\gamma, S_\gamma)}$$

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.

$$\frac{\partial \mathrm{Var}_{t+1}(S_\gamma)}{\partial \alpha}\Big|_{\mathrm{Var}_t(S_\gamma),\mathrm{Var}_t(G_\gamma),\mathrm{Cov}_t(G_\gamma,S_\gamma)} > 0 \tag{19}$$

2.4 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:

$$\text{Var}(G_\gamma) = 2\text{Var}(\epsilon)\left(1 + \frac{2w_G^2\text{Var}(\epsilon)}{w_G^2\left(\text{Var}(G_E) - 2\text{Var}(\epsilon)\right) + w_S^2\left(\text{Var}(S_E) - 2\text{Var}(\epsilon)\right)}\right) \tag{20}$$

$$\text{Var}(S_\gamma) = 2\text{Var}(\epsilon)\left(1 + \frac{2w_S^2\text{Var}(\epsilon)}{w_G^2\left(\text{Var}(G_E) - 2\text{Var}(\epsilon)\right) + w_S^2\left(\text{Var}(S_E) - 2\text{Var}(\epsilon)\right)}\right) \tag{21}$$

$$\text{Cov}(G_\gamma, S_\gamma) = 2\text{Var}(\epsilon)\left(\frac{2\text{Var}(\epsilon)w_G w_S}{w_G^2\left(\text{Var}(G_E) - 2\text{Var}(\epsilon)\right) + w_S^2\left(\text{Var}(S_E) - 2\text{Var}(\epsilon)\right)}\right) \tag{22}$$

To give more intuition, I'll define a few objects that will give the equations a much simpler and clearer form.

For comparison, consider random mating, which is defined as non-assortative mating. I'll use hats to denote values under random mating. The stationarity condition for $\text{Var}(G_\gamma)$ would be:

$$\begin{aligned}\widehat{\text{Var}(G_\gamma)} =& \text{Var}\left(\frac{1}{2}G_{\gamma,\text{Father}} + \frac{1}{2}G_{\gamma,\text{Mother}} + \epsilon_G\right) \\ =& \frac{1}{2}\text{Var}(G_\gamma) + \text{Var}(\epsilon_G) \\ =& 2\text{Var}(\epsilon_G) \\ =& 2\text{Var}(\epsilon)\end{aligned} \tag{23}$$

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[8] of general ability would be:

$$h^2(G) = \frac{\text{Var}(G_\gamma)}{\text{Var}(G)} = \frac{\text{Var}(G_\gamma)}{\text{Var}(G_\gamma) + \text{Var}(G_E)} \tag{24}$$

Under random mating, this would be:

$$\widehat{h^2(G)} = \frac{2\text{Var}(\epsilon)}{2\text{Var}(\epsilon) + \text{Var}(G_E)} \tag{25}$$

Using the heritability under random mating for general ability and system-

[8]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.

izing, I can rewrite the result as:

$$\text{Var}(G_\gamma) = \widehat{\text{Var}(G_\gamma)} \left(1 + \frac{w_G^2}{w_G^2 \left(\frac{1}{\widehat{h^2(G)}} - 2 \right) + w_S^2 \left(\frac{1}{\widehat{h^2(S)}} - 2 \right)} \right) \tag{26}$$

$$\text{Var}(S_\gamma) = \widehat{\text{Var}(S_\gamma)} \left(1 + \frac{w_S^2}{w_G^2 \left(\frac{1}{\widehat{h^2(G)}} - 2 \right) + W_w^2 \left(\frac{1}{\widehat{h^2(S)}} - 2 \right)} \right) \tag{27}$$

$$\text{Cov}(G_\gamma, S_\gamma) = 2\text{Var}(\epsilon) \left(\frac{w_G w_S}{w_G^2 \left(\frac{1}{\widehat{h^2(G)}} - 2 \right) + w_S^2 \left(\frac{1}{\widehat{h^2(S)}} - 2) \right)} \right) \tag{28}$$

The stationary variance under assortative mating is the the variance under 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, $\widehat{h^2(G)}$ and $\widehat{h^2(S)}$ must both be less than $\frac{1}{2}$, though $\lim_{\widehat{h^2(G)} \to \frac{1}{2}, \widehat{h^2(S)} \to \frac{1}{2}} \text{Var}(G_\gamma) = \infty$. 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 $\frac{1}{2}$.

2.5 Comparative Statics

I can get comparative statics results for the simplified model:

	w_G	w_S	$\text{Var}(G_E)$	$\text{Var}(S_E)$	$\text{Var}(\epsilon)$
$\text{Var}(G_\gamma)$	$+$	$-$	$-$	$-$	$+$
$\text{Var}(S_\gamma)$	$-$	$+$	$-$	$-$	$+$
$\text{Cov}(G_\gamma, S_\gamma)$	$+/-$	$+/-$	$-$	$-$	$+$

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 $w_G > w_S$.

	w_G	w_S	$\text{Var}(G_E)$	$\text{Var}(S_E)$	$\text{Var}(\epsilon)$
$\text{Var}(G_\gamma)$	$+$	$-$	$-$	$-$	$+$
$\text{Var}(S_\gamma)$	$-$	$+$	$-$	$-$	$+$
$\text{Cov}(G_\gamma, S_\gamma)$	$-$	$+$	$-$	$-$	$+$

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 6, 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 as 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),[9] 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.[10] I get a stronger result both by focusing on a trait which is heavily influenced by the parents' level 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.

[9] As a historical note, Fisher's paper is the first to use the word "variance" to denote the square of the standard deviation.

[10] Fernandez and Rogerson (2001) find a stronger effect than Kremer, though their approach is very different from both Kremer's and mine.

2.6 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 σ, I will show how much the area past T increases as I increase σ.

That is

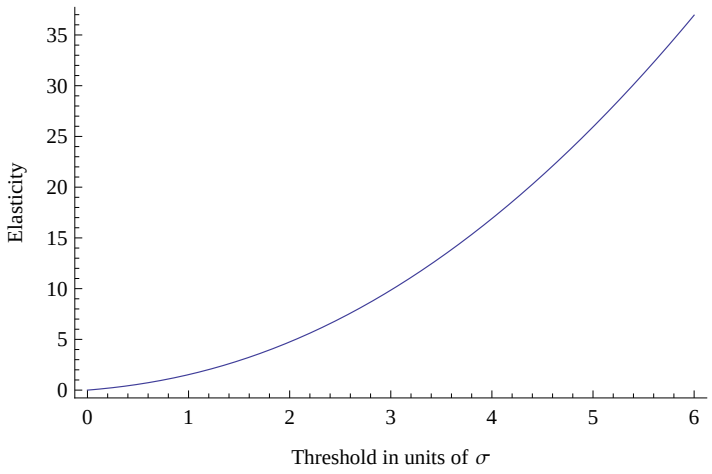
$$\begin{aligned}
\eta_\sigma(T,\sigma) :=& \frac{\sigma}{\text{Area}(T,\sigma)} \frac{\partial}{\partial \sigma} \left[\text{Area}(T,\sigma)\right] \\
=& \frac{\sigma}{1 - \frac{1}{\sqrt{2\pi}\sigma} \int_0^T e^{\frac{-x^2}{2\sigma^2}} dx} \frac{\partial}{\partial \sigma} \left[1 - \frac{1}{\sqrt{2\pi}\sigma} \int_0^T e^{\frac{-x^2}{2\sigma^2}} dx \right] \\
=& \sqrt{\frac{2}{\pi}} \frac{T}{\sigma} \frac{e^{\frac{-T^2}{2\sigma^2}}}{1 - \frac{2}{\sqrt{\pi}} \int_0^{\frac{T}{\sigma\sqrt{2}}} e^{-x^2} dx} \\
=& \eta_\sigma \left(\frac{T}{\sigma} \right)
\end{aligned} \tag{29}$$

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

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 σ 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 $\eta_\sigma(1) > 0$.

Below are numerical values for η_σ:



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.

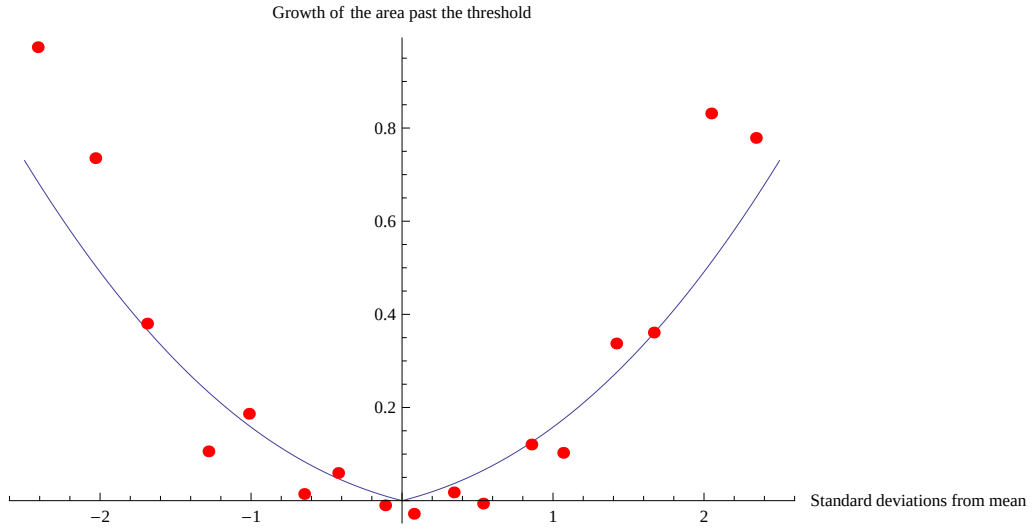
As another example, take the height distribution. The CDC runs a recurring study called the National Health and Nutrition Examination Survey (NHANES) which collects a variety of information, including height in a nationally representative sample, and versions of the study have existed since 1959.[11] This theory says that if the height distribution is approximately normal, 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

[11]The earliest study was called the National Health and Examination Survey (NHES). It is comparable for the variables I am interested in.

72.53 inches, which is 6.5% of the population:a 36% larger percentage than the percentage of the population taller than 72 inches in 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 $\eta_\sigma(1.67) = 3.47$, so our predicted increase would be $3.47 * 10.35\% = 35.9\%$ 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.

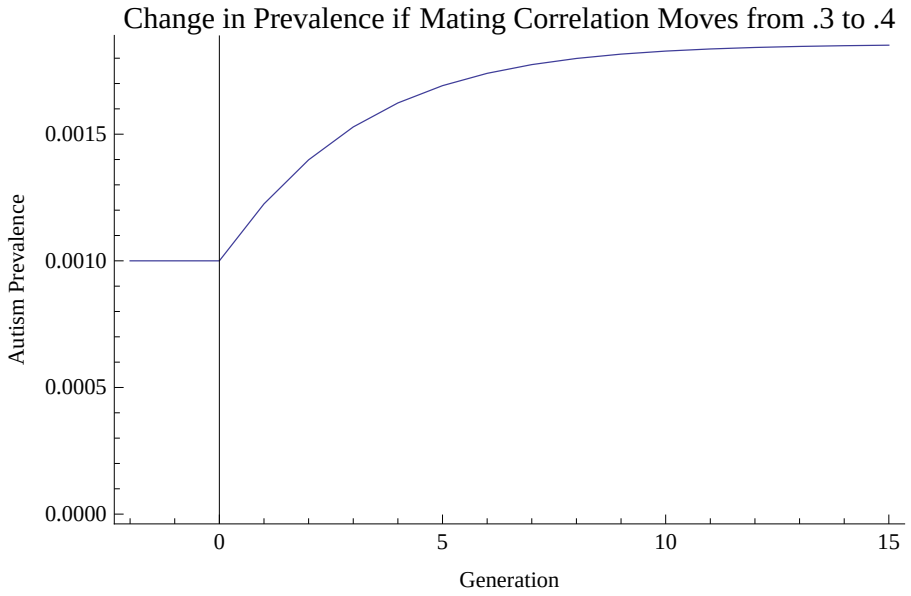


To be clear, this is not a deep test of the complete theory I have presented: all that is needed is for the height distribution to be normal in both 1959 and 2009. However, it is reassuring that the tails do grow at the high rates predicted, and the increase in the standard deviation is consistent with 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 very easy to explain in an assortative mating framework, but it is hard to explain through factors like nutrition that we typically discuss when thinking about long-term changes in the height distribution.

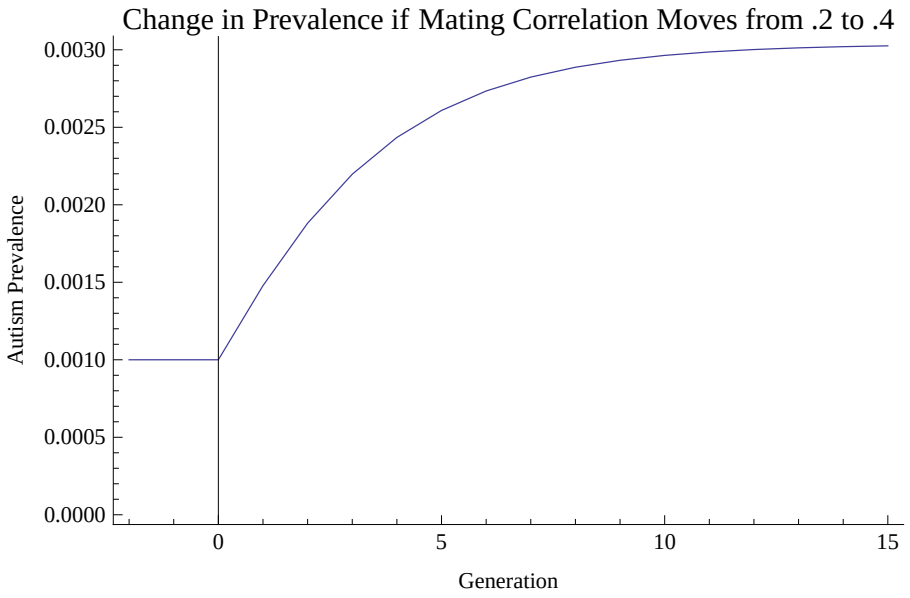
2.7 Potential Effect Size

Having seen in section 2.6 that relatively small changes in mating could have large effects on the extremes of distributions, I want to present a few examples 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 generaions 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.

3 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.[12] 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, individuals may become impaired because they are increasingly unable to deal with systems that do not have a lawful structure, such as social interactions and

[12]Baron-Cohen (2006)

verbal communication. Further, people at a very high systemizing level may become intensely interested in lawful, predictable things, like train schedules.[13] Those three features, social impairment, communications impairment, and restricted and repetitive behaviors and interests, are the three components of the diagnostic definition of autism.[14]

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.[15]

Baron-Cohen and co-authors have conducted a number of suggestive studies 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,[16] one showing a higher prevalence of engineers among the relatives of children with autism,[17] 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.[18]

Baron-Cohen also led a study (Roelfsema et al., 2011) which looked at the autism prevalence in three regions of the Netherlands: Eindhoven, Haarlem,

[13]Baron-Cohen (2006)

[14]DSM-IV-TR (2000)

[15]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.

[16]Baron-Cohen et al. (2007)

[17]Baron-Cohen et al. (1997)

[18]Baron-Cohen and Hammer (1997)

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, speciality providers, and all of the area's public schools.[19] Clinicians review the behavioral information collected in these records and make a determination, spending an average of 47 minutes on each child who does not have a previous clinical diagnosis.[20] 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.

[19] Rice et al. (2007)

[20] Van Naarden Braun et al. (2007)

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.

Mathematical ability is a good empirical analogue to systemizing ability. 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 combination of worker surveys and expert evaluation.[21] 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.[22] I include, as controls, the same measures that they used:

[21] 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.

[22] 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

median household income, the poverty rate, and the percent of adults with a bachelor's degree.[23] 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.

collaborative project with researchers at those sites.

[23]All measures come from the 2000 Census and all are also provided at the block group level. Though 2010 data are available, the block group boundaries have changed.

	(1)	(2)	(3)	(4)
	ASD	Cerebral Palsy	ASD	Cerebral Palsy
Avg Math Importance	0.264***	0.0164	0.289***	0.0634
	(0.0237)	(0.0260)	(0.0552)	(0.0596)
Poverty Rate			-0.128***	-0.110**
			(0.0307)	(0.0371)
% of Adults with BA			-0.138*	-0.101
			(0.0627)	(0.0625)
Med. Household Income			0.00298	-0.0496
			(0.0483)	(0.0501)
Observations	1,333	1,333	1,332	1,332
R-squared	0.070	0.000	0.084	0.011

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.

4 Rising rates of autism

While diagnosed rates of autism spectrum disorders have grown tremendously, there has been a great deal of debate about how much actual growth there has been, and some have even argued true prevalence hasn't risen much, or even at all.[24] In this section, I present suggestive evidence that ASD rates have truly risen.

Many claims that true ASD rates have not actually risen argue that diagnosis has broadened (e.g. King and Bearman (2009)), or that awareness has increased (e.g. Wing and Potter (2002)). While both of these effects have very likely contributed to the rise in diagnosed rates, neither are necessarily inconsistent with an actual rise. In this section, I'll focus on cases of autism where the children have an IQ below 70. I'll do this using the same data as section 3, as well as charts and tables from CDC reports which include data from other surveillance sites. An IQ below 70 is a key part of the criteria for a diagnosis of intellectual disability (ID), and is about two standard deviations below the mean IQ.[25] This is both a way of focusing on more severe cases of autism (the low IQ score is most likely caused by autism) and is a way of focusing on cases that are easier to detect. Cases of this severity don't fit with a story of increasingly marginal cases being labelled as autism, and eight year olds with a very low IQ are much more likely to have interacted with professionals who would be able to recognize signs of autism, so detection may be more consistent. At that very least, such children are very likely to have a

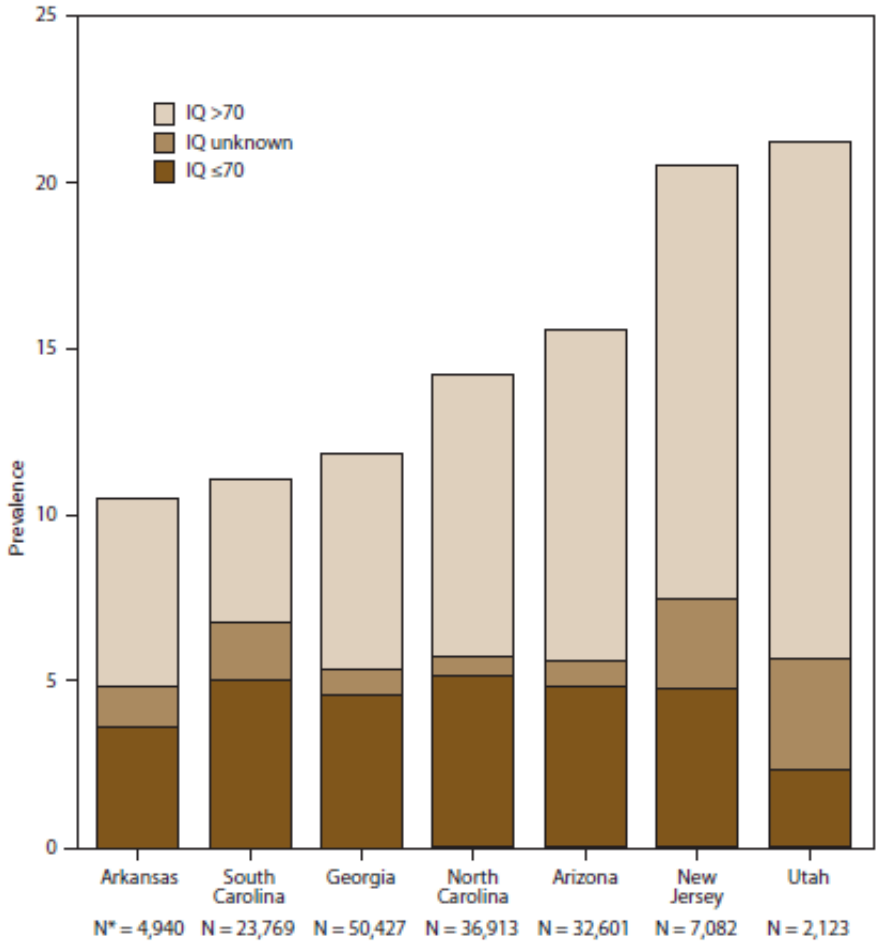
[24] See, for example, Gernsbacher, Dawson and Hill Goldsmith (2005)

[25] Intellectual disability is the disorder which used to be known as mental retardation.

diagnosis of intellectual disability, which is a fact I exploit in this section.

The data from section 3 comes from MADDSP which looks at eight-year olds in Metro Atlanta. MADDSP is part the Autism and Developmentmenal Disabilities Monitoring (ADDM) Network, which also includes sites elsewhere in the country. After every two-year surveillance cycle, a report is published which estimates national autism prevalence, and provides a variety of related statistics. The report issued in 2012 is Baio (2012).

When an ADDM site identifies an eight-year old as having an ASD, they try to obtain an IQ score from the child's records. Below is Baio (2012)'s Figure 4, which shows the prevalence of autism by site, broken down by IQ. The sites pictured as those for which an IQ could be obtained for at least 70% of children who were identified as having an ASD, and the sample sizes listed at the bottom are the population of eight-year olds in the surveillance area.



Besides the two sites with the smallest sample sizes (Arkansas, the first, and Utah, the last), the estimated prevalence among those with IQ≤70 is quite consistent, around 5 in 1000, while the prevalence among those with IQ>70 varies much more.[26] The variation in prevalence for higher IQ cases is consistent with the story that diagnostic criteria can vary for less severe cases. Even though we see a lot of site variation among the less severe cases, there's

[26]I would guess that unknown IQs are more likely to be above 70. Having an IQ below 70 is part of the DSM-IV-TR (2000) definition of intellectual disability, so intelligence testing is typically required before a child could be in special education for ID. If low IQ were suspected in a child, this gives an incentive to test for it.

a great deal more uniformity for the more severe cases, suggesting that these cases may be more consistently measured.

But even among these more severe cases, ASD rates have grown dramatically. Below are the estimated growth rates from Baio (2012):

Growth rates in ASD prevalence between 2002 and 2008

	All sites	MADDSP only
IQ$\leq$ 70	45.4%***	33.7%***
All IQs	78.5%***	58.3%***

*** p<0.01, ** p<0.05, * p<0.1
Estimates are from Baio (2012)

Significant growth among more severe cases is more difficult to explain through mechanisms like a broadening of diagnostic standards or awareness, and makes a stronger case for a true increase.[27] However, there is a version of the awareness argument that's worth investigating. It is possible that these more severe cases were detected in earlier study years, but that they were classified as simply having intellectual disability (ID), which requires an IQ of 70 or below, because clinicians didn't recognize signs of ASDs. When similar cases were detected in later years, they were classified as having ID and an ASD as awareness rose, leading to a apparent rise in more severe ASD cases even if the true prevalence didn't rise. Supporting that possibility, Croen et al. (2002) find that, for a set of cohorts in California, rates of ID without an ASD

[27]My calculations in section 2.7 would lead us to expect the lower IQ cases to be growing at a faster rate, since they're further along the tail, though it's likely that increased awareness and a broadening of disagnostic standards had a greater effect on the overall numbers, so this is not a clean comparison.

fell at about the same rate that ASD rates grew, suggesting that the growth of ASDs could be due just to diagnostic substitution.

Since I have ID and ASD cases by block group for 2000, 2002, 2004, 2006, and 2008 from MADDSP, I can look at this directly. We can measure how much ASD rates have risen across years, and look at how much ID rates have risen or fallen.

	(1)	(2)	(3)
	ID cases per 1000	ID cases per 1000	ASD cases per 1000
Year	0.313***	-0.0578	0.912***
	(0.0967)	(0.0933)	(0.0847)
ASD cases per 1000		0.407***	
		(0.0221)	
Observations	6,675	6,675	6,675
R-squared	0.002	0.123	0.019

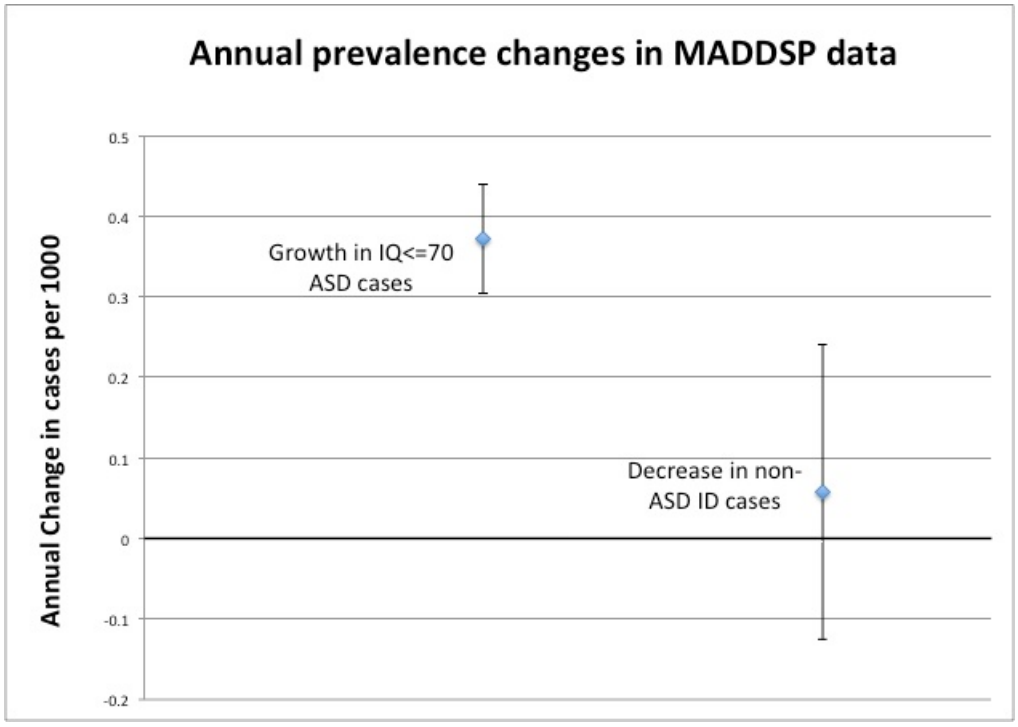
Robust standard errors in parentheses

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

As regressions 1 and 3 show, there's an upward trend in the rates of both ID and ASDs. In regression 2, the relationship between ASD rates and ID rates is mechanical, since it simply reflects that about 40% of ASD cases identified are children who also have an IQ at or below 70. Once we control for ASD cases (in regression 2), there is no upward trend in ID rates, and the point

estimate of the trend is negative though not significant.

Since, as we saw above, 40.7% of ASD cases identified by MADDSP involve children with an IQ below 70, we can compare the growth rate of these ASD cases to the decline in non-ASD ID cases. If we're only seeing diagnostic substitution, we should expect these to be equal.



This is consistent with a true increase in autism rates, since the low-IQ ASD cases are rising faster than non-ASD ID cases are falling, suggesting that we're not simply seeing diagnostic substitution.

5 Assortative Mating

If autism is the extreme right tail of the systemizing distribution, section 2 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,[28] 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.[29] However, an increase in marriages 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.

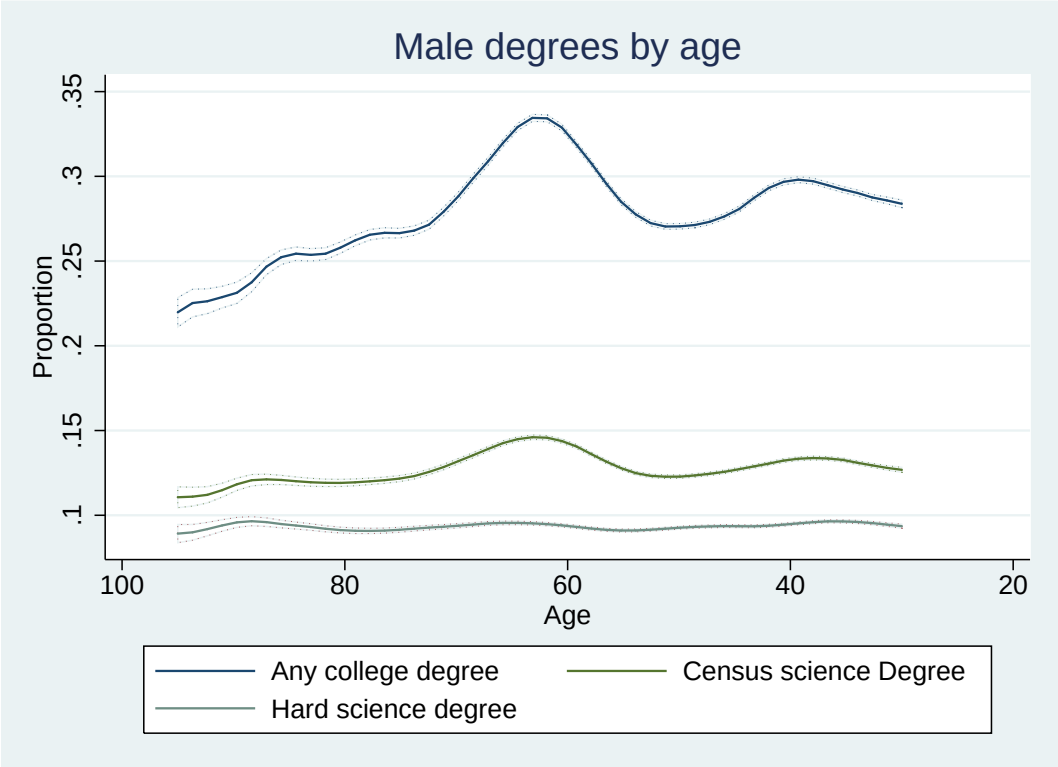
5.1 Matching on Degrees

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 definition includes degrees in

[28] e.g. Baron-Cohen (2006)
[29] Silberman (2001)

psychology and in the social sciences (both of which are specific categories). Because these categories may not involve the same level of systemizing, I'll also use a narrow measure which excludes those two ("hard science")[30].

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).

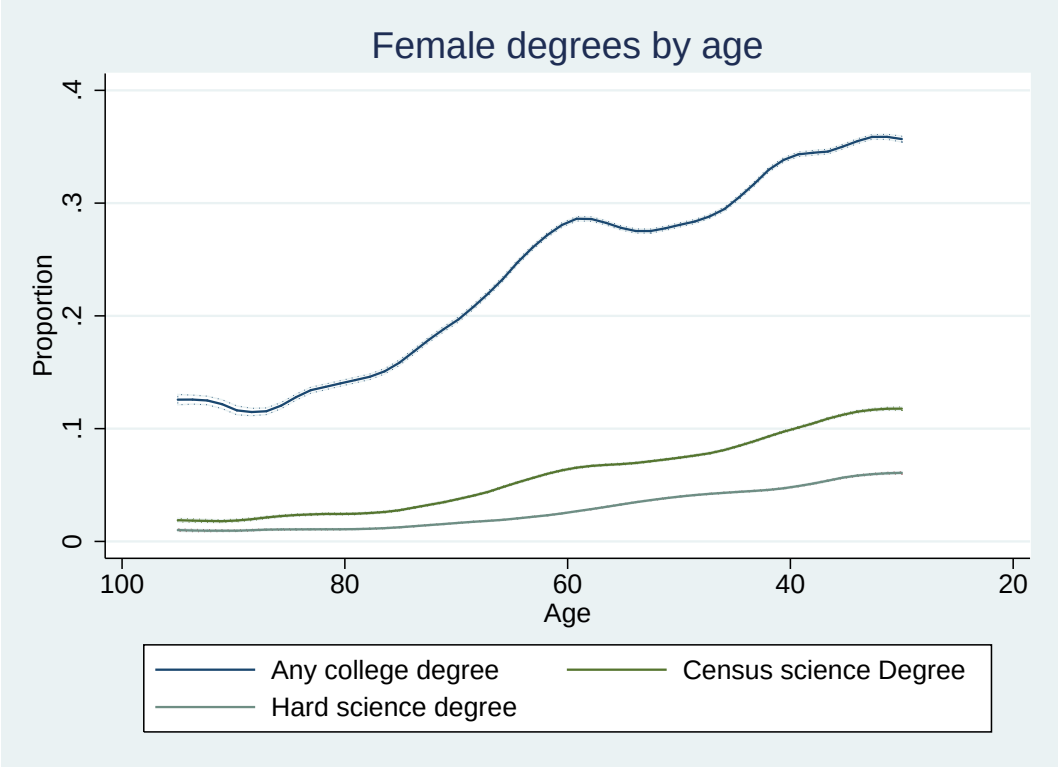


There are two things to observe in the graph. First, the graph supports the idea that field of degree is a meaningful measure. The spike and then fall in degree attainment that begins around age 60 is driven by the increasing and then decreasing returns to college in the 1970's.[31] The spike is much

[30]The specific categories that remain are computers and math, biological sciences, physical science, engineering, and multidisciplinary studies.

[31]Katz and Murphy (1992)

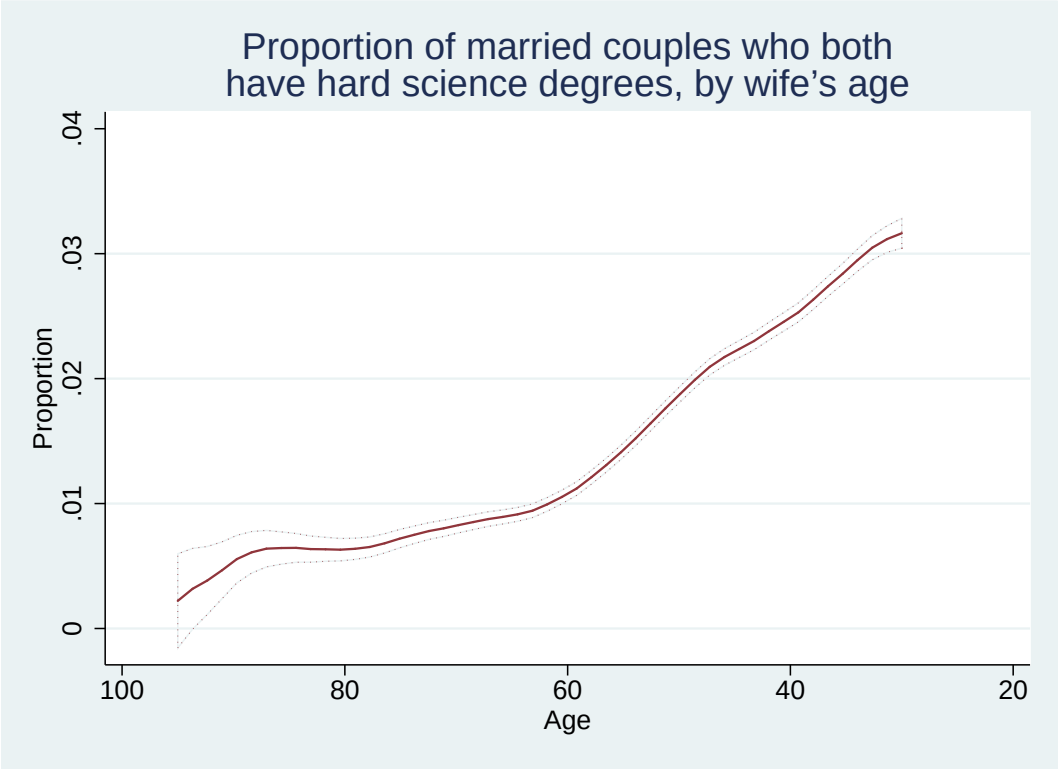
more muted when we look at census science degrees, and is basically absent when we look at hard science degrees. If field choice were purely random, we would see a proportional spike in each line. The absence of a spike in hard science degrees suggests that these require certain abilities that are not common among those on the margin for college. Second, the portion of men with hard science degrees has remained nearly constant, at just under 10%.



Women experienced a steep rise in the chance 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.[32]

[32]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.

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.



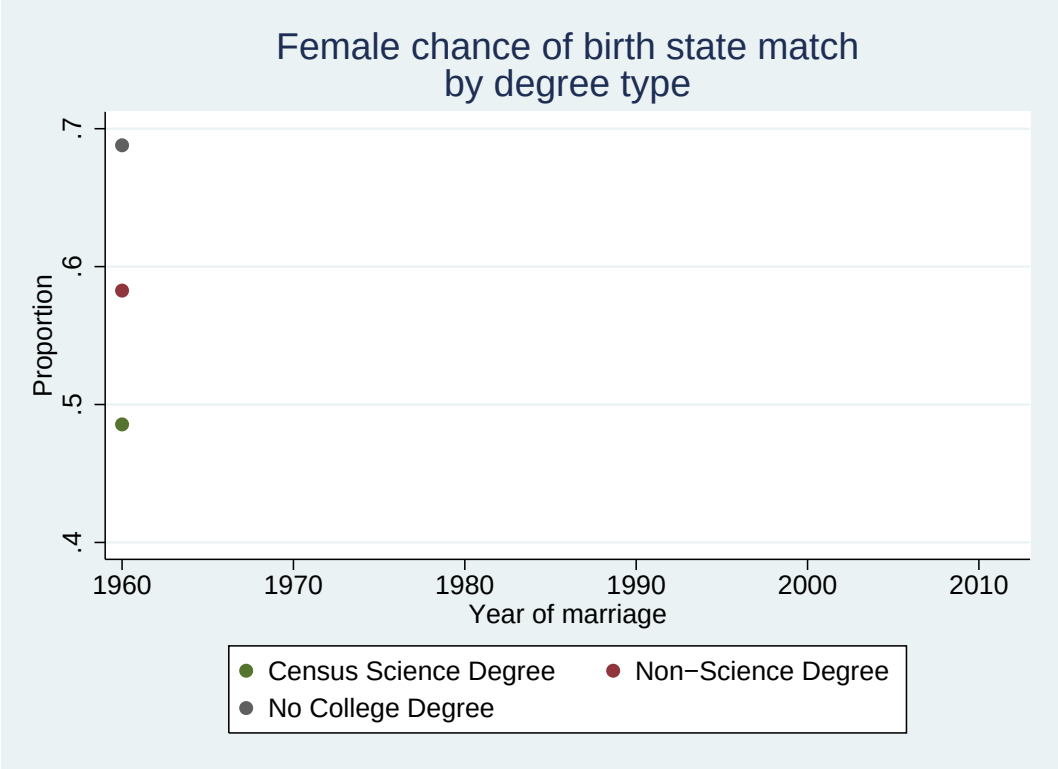
5.2 Matching Across States

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. In other words, the rapid growth of women with science degrees is not being driven by more women being qualified to get science degrees, so there are many older women who did not get science degrees, but would have if they'd been born later. The question is, would those women

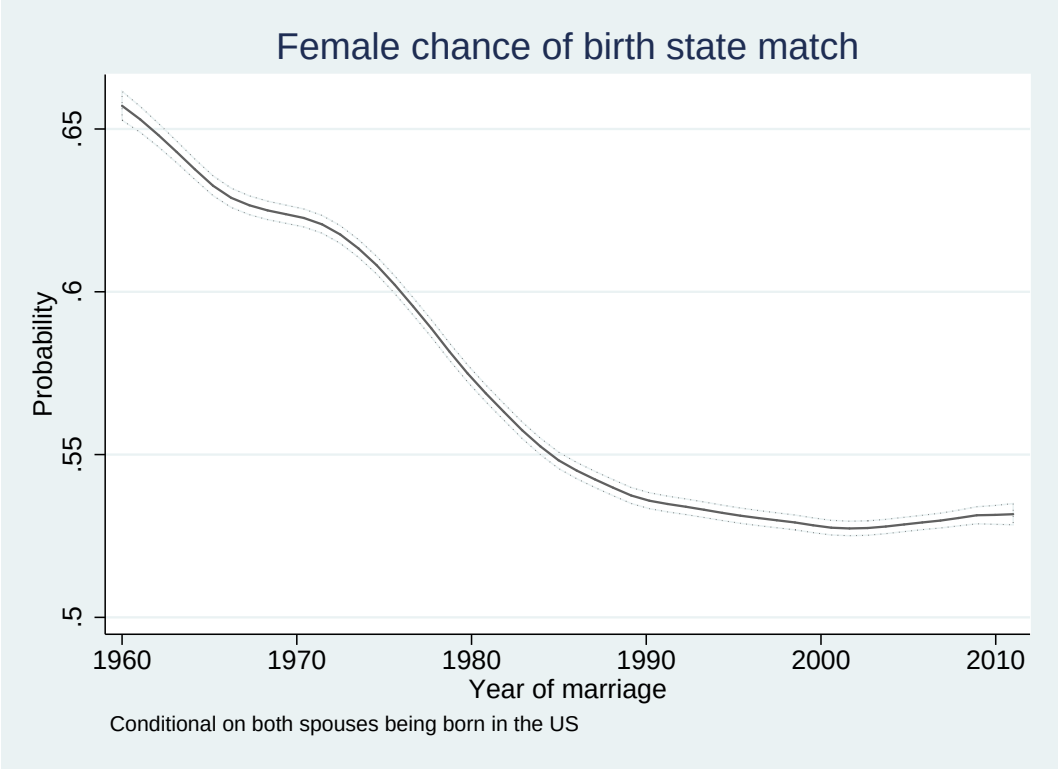
have married different men if they'd gotten science degrees, or would they have married the same men because what matters is ability and not degree attainment. If latent ability is all that matters, the rise is sorting on observables would just be an artifact. However, if degree attainment actually changes behavior, then the rise in sorting on observables is evidence of a meaningful change in sorting.

To get at this question, we can look at the tendency of people to marry someone born in their birth state. This is a marriage outcome that depends on human capital, and is something that we can observe consistently over time. I'll argue below that we would expect very different patterns in the data if latent ability drives birth state matching, versus if actual degree attainment drives birth state matching. The data seem most consistent with degree attainment being the driving factor, which suggests that a meaningful change in sorting has occurred.

Here are the probabilities that a women who married in 1960 married someone born in her birth state, by her educational attainment.



As you can see, the probability of a birth state match depends on 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.

Parameter	**Value**
Share of women with no degree in 1960-62 ($\alpha_{\text{No Degree}}$)	.814
Share of women with no degree in 2005-07 ($\alpha'_{\text{No Degree}}$)	.616
Share of women with a non-science degree in 1960-62 ($\alpha_{\text{Non-Sci}}$)	.159
Share of women with a non-science degree in 2005-07 ($\alpha'_{\text{Non-Sci}}$)	.274
Share of women with a science degree in 1960-62 (α_{Science})	.026
Share of women with a science degree in 2005-07 (α'_{Science})	.11

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

Degree type	**Initial birth state match probability**
No degree ($\beta_{\text{No Degree}}$)	.667
Non-science degree ($\beta_{\text{Non-Sci}}$)	.561
Science degree (β_{Science})	.514

The goal is to recover $\gamma_{\text{No Degree}}$, $\gamma_{\text{Non-Sci}}$, and γ_{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 β'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 η. So:

$$\beta_{\text{Science}} = \gamma_{\text{Science}}$$

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

$$\begin{aligned}\beta_{\text{No Degree}} = &\frac{\alpha'_{\text{No Degree}}\gamma_{\text{No Degree}} + (1-\eta)(\alpha'_{\text{Science}} - \alpha_{\text{Science}})\gamma_{\text{Science}}}{\alpha_{\text{No Degree}}} \\ &+ \frac{(\alpha'_{\text{Non-Sci}} - \alpha_{\text{Non-Sci}} + \eta\alpha'_{\text{Science}} - \alpha_{\text{Science}}))\gamma_{\text{Non-Sci}}}{\alpha_{\text{No Degree}}}\end{aligned}$$

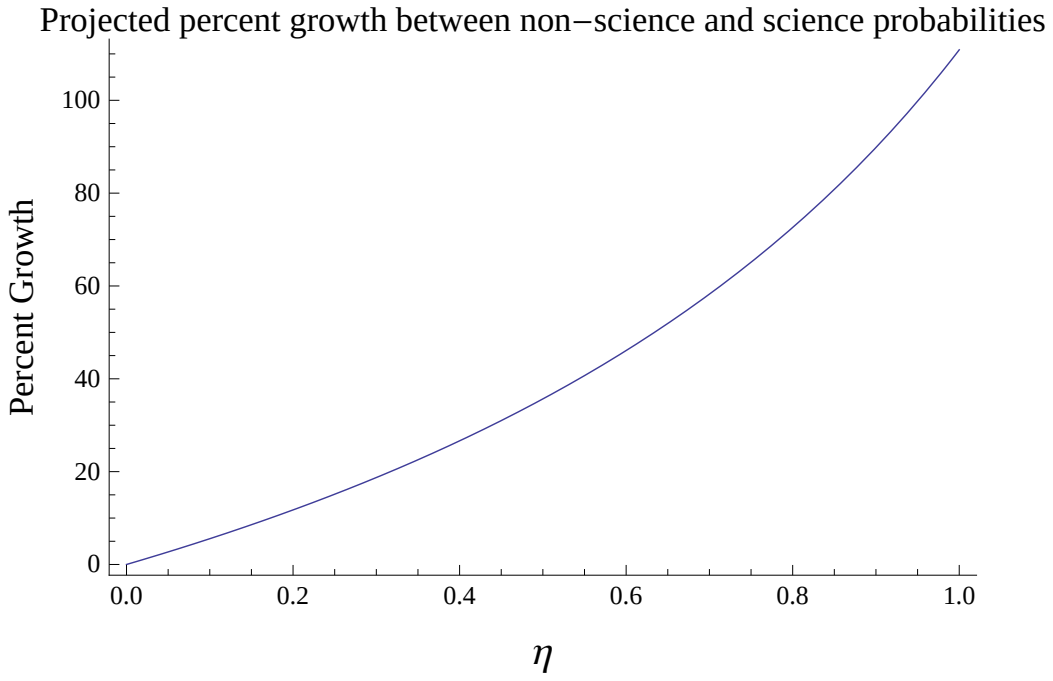
These equations can easily solved for the γ's. Because we have let η vary,

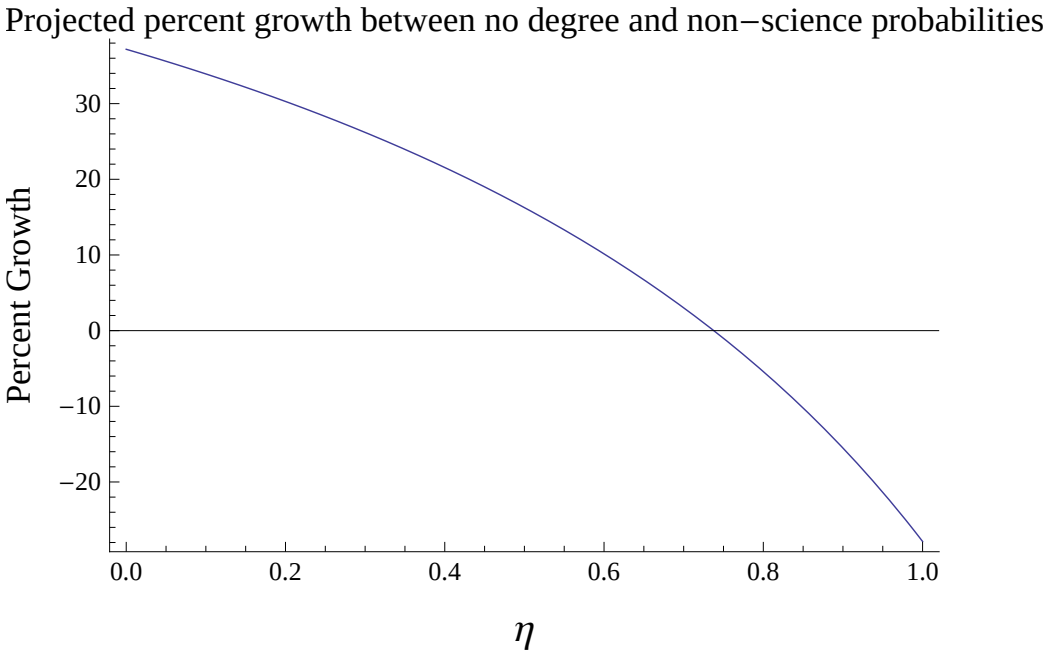
we will have $\gamma_{\text{Science}}(\eta)$, $\gamma_{\text{Non-Sci}}(\eta)$, and $\gamma_{\text{No Degree}}(\eta)$. Since the probability for science degrees does not change, the percent growth in the distance between the science and non-science probabilities would be:

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

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

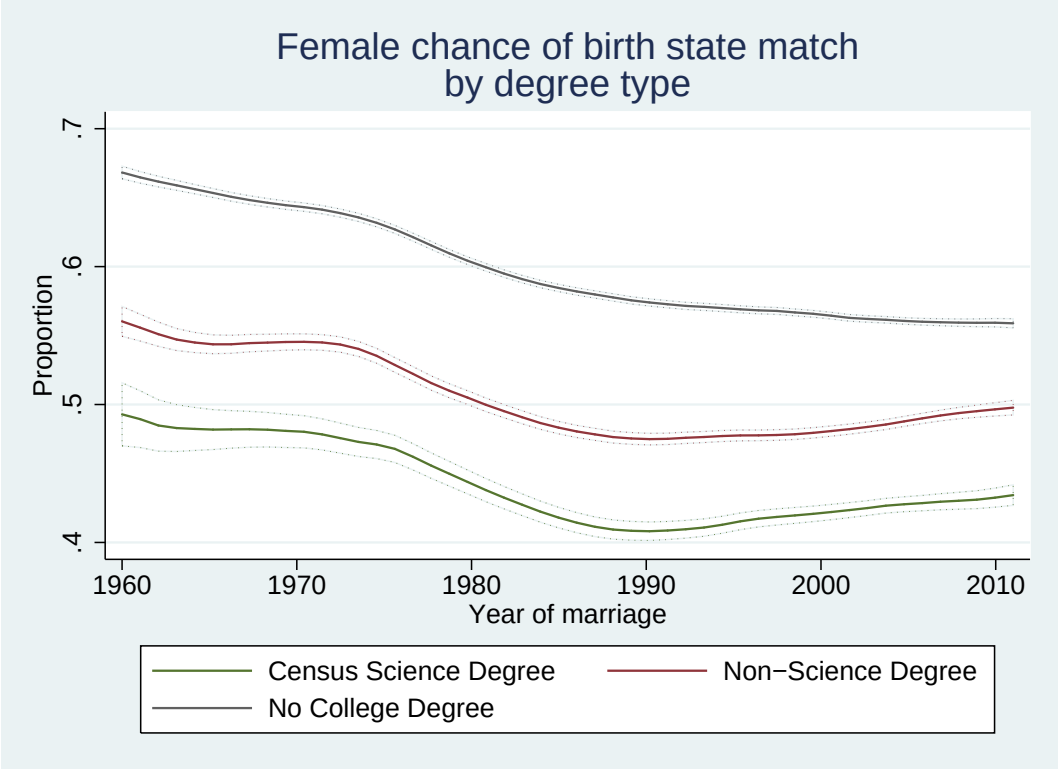
$$\frac{\gamma_{\text{No Degree}}(\eta) - \beta_{\text{No Degree}} - (\gamma_{\text{Non-Sci}}(\eta) - \beta_{\text{Non-Sci}})}{\beta_{\text{No Degree}} - \beta_{\text{Non-Sci}}}$$





As you can see from the projected charts, if latent ability mattered, the lines should have moved noticeably apart. At very low values of η, we would have seen significant growth between the no degree and non-science lines. At any other level of η, 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.[33]

6 Welfare and Policy

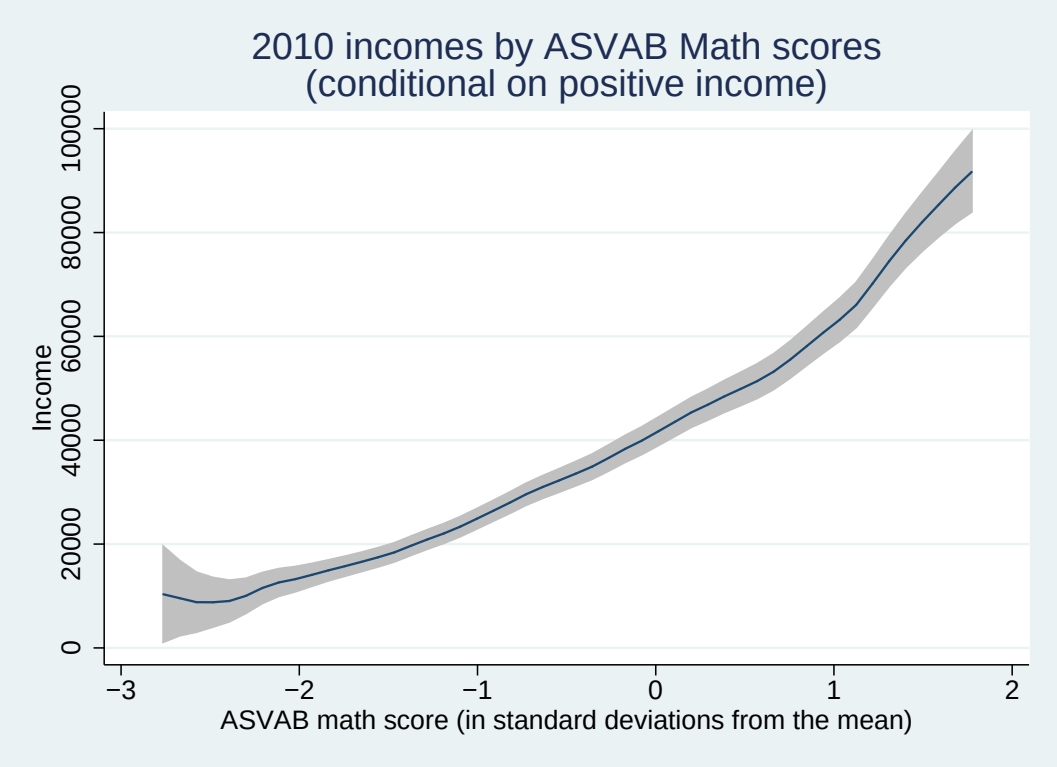
One response to the idea that increased assortative mating on systemizing has increased the autistic population would be to discourage systemizers from marrying. However, I believe that this would be misguided. Despite the high

[33]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.

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 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 NLSY 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.[34] Below are the smoothed means:

[34]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:

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

I will also need to estimate costs along the autistic spectrum. Ganz (2007) estimated the lifetime social costs of a case of autism, including medical and

non-medical care, and lost productivity, and arrived at $3.2 million.[35] To be generous, let's call it $5 million.[36] 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%.

[35]Ganz (2007)

[36]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

$$\text{Net benefit} = \underbrace{30 \int_{-\infty}^{0} \left(\hat{\beta}_{\text{Below}} x + \hat{\gamma}_{\text{Below}} x^2\right) \left(\frac{1}{1.1\sqrt{2\pi}} \mathrm{e}^{\frac{x^2}{2(1.1)^2}} - \frac{1}{1\sqrt{2\pi}} \mathrm{e}^{\frac{x^2}{2(1)^2}}\right) \mathrm{d}x}_{\text{Lower lifetime earnings on the bottom half of the distribution}}$$

$$+ \underbrace{30 \int_{0}^{2.5} \left(\hat{\beta}_{\text{Above}} x + \hat{\gamma}_{\text{Above}} x^2\right) \left(\frac{1}{1.1\sqrt{2\pi}} \mathrm{e}^{\frac{x^2}{2(1.1)^2}} - \frac{1}{1\sqrt{2\pi}} \mathrm{e}^{\frac{x^2}{2(1)^2}}\right) \mathrm{d}x}_{\text{Higher lifetime earnings on the top half of the distribution}}$$

$$+ \underbrace{\int_{2.5}^{3} 0 \left(\frac{1}{1.1\sqrt{2\pi}} \mathrm{e}^{\frac{x^2}{2(1.1)^2}} - \frac{1}{1\sqrt{2\pi}} \mathrm{e}^{\frac{x^2}{2(1)^2}}\right) \mathrm{d}x}_{\text{Asperger's is net neutral}}$$

$$+ \underbrace{\int_{3}^{\infty} -5,000,000 \left(\frac{1}{1.1\sqrt{2\pi}} \mathrm{e}^{\frac{x^2}{2(1.1)^2}} - \frac{1}{1\sqrt{2\pi}} \mathrm{e}^{\frac{x^2}{2(1)^2}}\right) \mathrm{d}x}_{\text{The lifetime cost of each new case of autism}}$$

$$= \underbrace{-\$21,869}_{\text{Lower lifetime earnings on the bottom half of the distribution}}$$

$$+ \underbrace{\$42,217}_{\text{Higher lifetime earnings on the top half of the distribution}}$$

$$+ \underbrace{\$0}_{\text{Asperger's is net neutral}}$$

$$+ \underbrace{-\$12,896}_{\text{The lifetime cost of each new case of autism}}$$

$$= \$11,132$$

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, on balance, increased assortative mating on systemizing is probably a good thing.

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.

7 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 such as computers, which have likely increased the returns to systemizing, have led to more assortative mating on systemzing. 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 be playing 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

American Psychiatric Association. 2000. *Diagnostic and statistical manual of mental disorders.* . 4th ed., tex rev. ed., United States:American Psychiatric Press Inc (DC).

Baio, Jon. 2012. "Prevalence of Autism Spectrum Disorders: Autism and Developmental Disabilities Monitoring Network, 14 Sites, United States, 2008." *Morbidity and Mortality Weekly Report*, 61(SS-3): 1–19.

Baron-Cohen, Simon. 2006. "The hyper-systemizing, assortative mating theory of autism." *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, 30(5): 865872.

Baron-Cohen, Simon, and Jessica Hammer. 1997. "Parents of children with Asperger Syndrome: What is the cognitive phenotype?" *Journal of Cognitive Neuroscience*, 9(4): 548–554.

Baron-Cohen, Simon, Sally Wheelwright, Amy Burtenshaw, and Esther Hobson. 2007. "Mathematical Talent is Linked to Autism." *Human Nature*, 18(2): 125–131.

Baron-Cohen, Simon, Sally Wheelwright, Carol Stott, Patrick Bolton, and Ian Goodyer. 1997. "Is there a link between engineering and autism?" *Autism*, 1(1): 101109.

Becker, Gary S, and Nigel Tomes. 1979. "An equilibrium theory of the distribution of income and intergenerational mobility." *The Journal of Political Economy*, 87(6): 11531189.

Bulmer, Michael. 1980. *The mathematical theory of quantitative genetics.* Oxford:Clarendon Pr.

Croen, Lisa A., Judith K. Grether, Jenny Hoogstrate, and Steve Selvin. 2002. "The Changing Prevalence of Autism in California." *Journal of Autism and Developmental Disorders*, 32(3): 207–215.

Durkin, Maureen S., Matthew J. Maenner, F. John Meaney, Susan E. Levy, Carolyn DiGuiseppi, Joyce S. Nicholas, Russell S. Kirby, Jennifer A. Pinto-Martin, and Laura A. Schieve. 2010. "Socioeconomic Inequality in the Prevalence of Autism Spectrum Disorder: Evidence from a U.S. Cross-Sectional Study." *PLoS ONE*, 5(7): e11551.

Fernandez, Raquel, and Richard Rogerson. 2001. "Sorting and Long-Run Inequality." *Quarterly Journal of Economics*, 116(4): 13051341.

Fisher, Ronald A. 1918. "The correlation between relatives on the supposition of Mendelian inheritance." *Transactions of the Royal Society of Edinburgh*, 52: 399–433.

Freitag, Christine M. 2006. "The genetics of autistic disorders and its clinical relevance: a review of the literature." *Molecular Psychiatry*, 12(1): 222.

Ganz, Michael L. 2007. "The lifetime distribution of the incremental societal costs of autism." *Archives of Pediatrics & Adolescent Medicine*, 161(4): 343–349.

Gernsbacher, M. A, M. Dawson, and H. Hill Goldsmith. 2005. "Three reasons not to believe in an autism epidemic." *Current Directions in Psychological Science*, 14(2): 55.

Katz, Lawrence F., and Kevin M. Murphy. 1992. "Changes in Relative Wages, 1963-1987: Supply and Demand Factors." *The Quarterly Journal of Economics*, 107(1): 35–78.

King, Marissa, and Peter Bearman. 2009. "Diagnostic change and the increased prevalence of autism." *International Journal of Epidemiology*, 38(5): 1224–1234.

Kremer, Michael. 1997. "How Much Does Sorting Increase Inequality?*." *Quarterly Journal of Economics*, 112(1): 115139.

Newschaffer, Craig J., Lisa A. Croen, Julie Daniels, Ellen Giarelli, Judith K. Grether, Susan E. Levy, David S. Mandell, Lisa A. Miller, Jennifer Pinto-Martin, Judy Reaven, Ann M. Reynolds, Catherine E. Rice, Diana Schendel, and Gayle C. Windham. 2007. "The Epidemiology of Autism Spectrum Disorders." *Annual Review of Public Health*, 28(1): 235–258.

Rice, Catherine E., Jon Baio, Kim Van Naarden Braun, Nancy Doernberg, F. John Meaney, and Russell S. Kirby. 2007. "A public health collaboration for the surveillance of autism spectrum disorders." *Paediatric and Perinatal Epidemiology*, 21(2): 179190.

Roelfsema, Martine T., Rosa A. Hoekstra, Carrie Allison, Sally Wheelwright, Carol Brayne, Fiona E. Matthews, and Simon Baron-Cohen. 2011. "Are Autism Spectrum Conditions More Prevalent in an Information-Technology Region? A School-Based Study of Three Regions in the Netherlands." *Journal of Autism and Developmental Disorders*, 42(5): 734–739.

Silberman, Steve. 2001. "The Geek Syndrome." *Wired Magazine*, , (12).

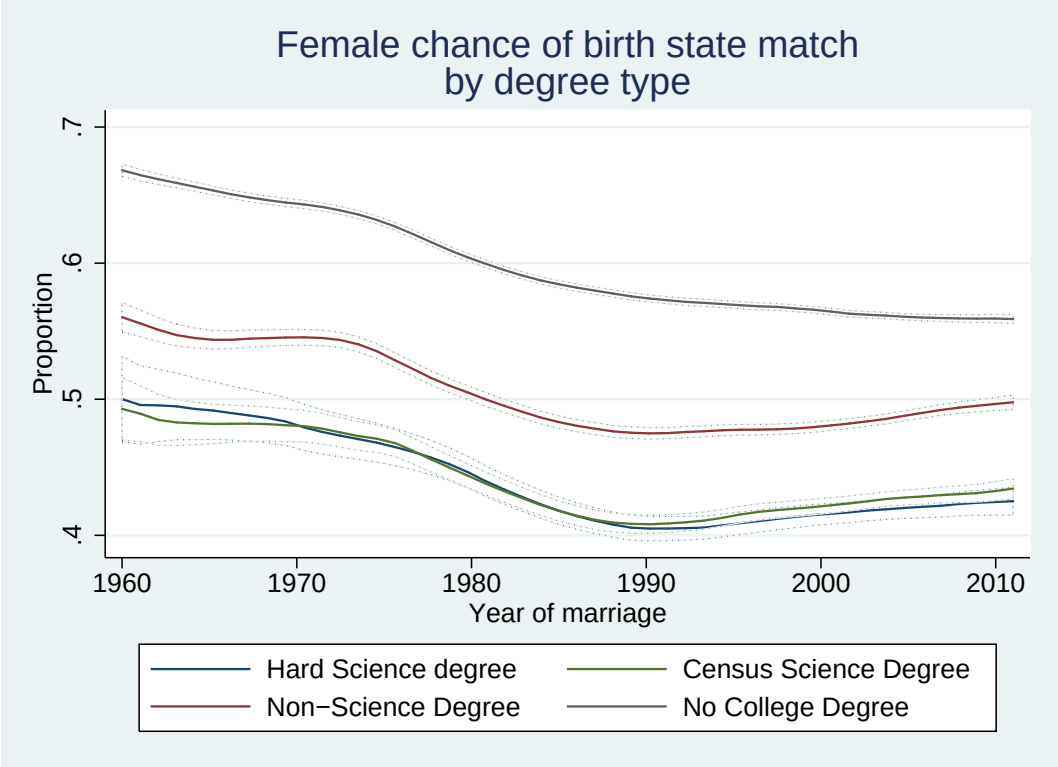
Thiessen, Del, and Barbara Gregg. 1980. "Human assortative mating and genetic equilibrium: An evolutionary perspective." *Ethology and Sociobiology*, 1(2): 111–140.

Van Naarden Braun, Kim, Sydney Pettygrove, Julie Daniels, Lisa Miller, Joyce Nicholas, Jon Baio, Laura Schieve, Russell S. Kirby, Anita Washington, Sally Brocksen, Hossein Rahbar, and Catherine Rice. 2007. "Evaluation of a Methodology for a Collaborative Multiple Source Surveillance Network for Autism Spectrum Disorders — Autism and Developmental Disabilities Monitoring Network, 14 Sites, United States, 2002." *MMWR Morb Mortal Wkly Rep*, 56(SS01): 29–40.

Wing, Lorna, and David Potter. 2002. "The epidemiology of autistic spectrum disorders: is the prevalence rising?" *Mental Retardation and Developmental Disabilities Research Reviews*, 8(3): 151161.

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 5.2. Since the probabilities are so similar, the additional structure would not significantly alter the analysis.



B Appendix B: Regressions for Section 6

Below are the regressions for the rough welfare calculation in section 6. 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).

	(1)	(2)
	Income if Math>mean	Income if Math<mean
ASVAB Math	15,036**	-19,234
	(6,155)	(12,493)
ASVABMath^2	-1,230	32,207***
	(3,184)	(6,951)
Constant	48,771***	58,133***
	(2,496)	(4,653)
Observations	1,432	1,781
R-squared	0.034	0.080

ASVAB Math scores are in standard deviation units

Standard errors in parentheses

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