Steven: Hello, and welcome to the Inequality Podcast. I'm Steven Durlauf. Today, I'm delighted to introduce two guests who will be discussing important issues associated with gender inequality. It is natural in thinking about any contemporary inequalities to focus on gender. There are really universal differences in the socioeconomic status of men and women that has been true historically, true contemporaneously, and certainly true across many different societies-though the degrees of these inequalities, of course, vary profoundly. In thinking about the United States, there have been very profound and important improvements with reference to gender inequality, but disparities still exist, and there are complicated explanations and issues as to why they are so. Today's guests have made essential contributions to our understanding of gender inequality in contemporary life. Our first guest is Martha Bailey. She's a professor of economics at UCLA and the director of the California Center for Population Research at UCLA. One of the great empirical economists and economic historians of her generation, Martha's work examines the effects of contraception on male-female inequalities in the United States. She's also done important work understanding how anti-discrimination policies have affected male-female inequality and worked on the consequences of the War on Poverty. Later, you'll hear a conversation with my co-host Geoff Wodtke, with Natasha Quadlin, who is an associate professor of sociology at UCLA and a faculty fellow at the California Center for Population Research. They will discuss her research examining the unique challenges women college graduates face when they enter the labor market. But first, here is my conversation with Martha Bailey. I thought we would start with some of your recent work, which has looked at the effects of seminal War on Poverty legislation from the perspective of group inequality-in particular, gender inequality-the Equal Pay Act, the Civil Rights Act, and their implications for the current gender wage gap.

Martha: The Equal Pay Act was passed in 1963, which was before I would date the beginning of the War on Poverty, and the Civil Rights Act was passed the next year in 1964, which ended up including women in a somewhat last-minute and controversial addition. So this is one of my most recent papers on the 1960s. It was motivated by some work I was doing in the archives around the 1966 Fair Labor Standards Act. As I was digging around in these reports by the Department of Labor trying to understand the discussion around the amendment and its passage, I kept finding references to how successful the Equal Pay Act had been. I was really puzzled by this because I had been teaching in my undergraduate and graduate classes, based on the work of prominent scholars like recent Nobel Laureate Claudia Goldin and Francine Blau, that the Equal Pay Act had not been particularly effective in reducing sex discrimination, and neither had the Civil Rights Act's provisions around sex discrimination. This has been part of the consensus view for a very long time. The main reference there is the stability of the gender wage gap in the 1960s and 70s. So after these two important pieces of legislation were passed, there's really not much that goes

on in the gender wage gap—a lot of stability until we get to the 1980s. There are lots of reasons why people have concluded this legislation wasn't particularly impactful. The Equal Pay provision is so narrow that firms could comply by just changing job titles or a few details about the work description. The Civil Rights Act was really focused on racial discrimination, not sex discrimination, in the 1960s, and there wasn't a single case for sex discrimination litigated until the 1970s. The consensus view that these things didn't really do a lot until we get into the 1970s and the courts started intervening to enforce the Civil Rights Act has held for a very long time.

Martha: So when I was reading these Labor Department reports, I got really curious about this and was trying to figure out—often as a historian, when you read something that doesn't gel with your understanding of the world, you think, what's going on there? So after we wrapped up the minimum wage project, we started looking into this and brainstorming ways to test how the Equal Pay Act and Civil Rights Act could have mattered. This was together with two of my co-authors, Brian Stewart and Thomas Helgerman. The difficulty in studying these two pieces of legislation is actually similar to the minimum wage literature because a federal minimum wage hike just passed nationally and affects labor markets nationally. One way you might look at whether or not the legislation had an effect is to just look at changes over time. But the problem with that research design is that everything else going on can be conflated with the legislation, and trying to disentangle that is pretty complicated. That's the same thing we'd been doing with the gender wage gap, we've been leaning on this time series of the gender wage gap. So we decided to take a cue from the minimum wage literature and think about how the incidence of the Equal Pay Act or the Civil Rights Act and their implications for sex discrimination could really vary across space and across jobs. How could we document variation in how much it mattered? One of the things we came up with is the gender gap in 1960. The idea is that the gender gap doesn't just reflect pay discrimination in labor markets; it reflects a whole lot of other things too. But we thought, what if the gender gap is correlated with pay discrimination in a particular industry or occupation? So if there's more pay discrimination, we would expect a bigger gender gap in those areas. If the gender gap is correlated with pay discrimination against women and the legislation reduces it, then we should expect women's wages to grow more in those industries and occupations after 1964, when the legislation was passed, in jobs with larger pre-existing gaps. And it turns out that's exactly what comes out of the data. Candidly, I thought this was a crazy idea, but we decided to try it, and it worked. We find that women's wages grew much more quickly after 1964 in jobs where there were bigger gender gaps beforehand, and we have very nice parallel trends before the policy change took place. The biggest worry is that, for some reason, there was a large demand shock that happened in those industries and occupations where women were working, and it just happened to occur after the 1963 and 1964 legislation took effect. So

the other thing we did is look at whether we see similar patterns for men's wages. If there was a boom in those industries and particular types of jobs, we would expect to see men's wages increase in the aftermath as well. But in fact, we can reject that with the data. There's no evidence of positive gains for me;if anything, we see a slight decrease for them, which is consistent with the legislation raising the cost of labor more generally and potentially firms hiring fewer workers, men included.

Steven: Is there any exploration of differences by race?

Martha: We did look at differences by race, and this turns out to be really hard because Black women are a very small share of the sample in the Current Population Survey. We do see similar effect—wage gains for both Black women and White women—and we can't reject that they're the same. The magnitudes are roughly similar for the two groups. We don't see any effects for Black men. We're trying to rule out the effect of the Civil Rights Act operating through reductions in racial discrimination, which we think it did, but the idea is that that's not going to be correlated with the size of the gender gap we're measuring in jobs and occupations.

Steven: I thought you were a little too modest in the description of the empirical strategy. It's an extremely important and insightful idea to say that identification of the effects of policy on discrimination is going to require heterogeneities that you can exploit to implicitly differ out other factors. Most studies think about time as the dimension of heterogeneity. You're increasing the dimensionality by thinking about occupations and regions. That seems to me a very important lesson in general for the methods of empirical work on discrimination.

Martha: One of the things when we presented this—and I won't tell you which economist gave us this comment—is that the identification strategy just wasn't credible, and we needed another research design. So we doubled down and dug back through the literature and found another one. It turns out there were actually equal pay acts at the state level going all the way back to World War I, in some cases, but mostly passed after World War II. And despite what a lot of people might think—were these really about equality in labor markets for women? The answer is no. They were passed because states were worried that firms would hire lower-cost women instead of bringing back veterans. States with a lot of manufacturing have especially passed these equal pay acts. A lot of the arguments for the federal legislation are that they hadn't been particularly effective or were not particularly well enforced; that was one of the arguments the Labor Department made to Congress. The second identification strategy was: if those state equal pay laws are not perfectly effective but somewhat effective—more effective than not having any law—and we have two sets of states, those with pre-existing equal pay laws and those

without, we'd expect women's wages to grow more quickly in the states without the pre-existing equal pay laws, presuming that some of the adjustment had already happened in the states that had passed those laws much earlier. And again, that's exactly what we find. Neither piece of evidence is perfect, but we're doing the best we can to get our hands around the effects of the legislation. One of the hardest parts of this paper, and one of the questions I get asked the most, is: how could the legislation be so effective when you also start with this picture of the gender wage gap, where you don't see any change in the time series? So there are a couple of things we did to try to reconcile this. One of the analyses we did found that women's wage gains were really concentrated in lower-earning jobs. We think this is for several reasons, but one of them is because the Labor Department was enforcing the minimum wage, and those investigations were much more focused on certain types of jobs, especially among lower earners. There would have been a lot more "voluntary compliance" by firms in the places where they thought the Labor Department would go and look. The Equal Pay Act was an amendment to the Fair Labor Standards Act, so the two were bound together. When the Labor Department would go looking for violations of the Fair Labor Standards Act, they would also examine compliance with the Equal Pay Act. So those jobs, especially the lower-earning ones, fell under a microscope. The other piece is that the equality of work—that very narrow legal standard—was much easier to see for things like cashiers or people working at retail counters. You could see a man and a woman working at the same retail counter and one being paid two-thirds of what the other is being paid. That was something much easier to remedy than softer skills. How productive are you at an advertising agency? Have you seen Mad Men?

Steven: Yes, I have.

Martha: So Peggy, who works in advertising, and there's a part where she's talking about how she's gotten nothing from the Equal Pay Act. I think she was a fairly high earner right there, and that's consistent with our results. How do you compare Peggy's output to Don Draper's? That's much harder to do. So one of the reasons why women's wage gains were concentrated in lower-earning jobs after the Equal Pay Act was probably because it was easier to judge the equality of work. The second thing that I think is really interesting—and it's always hard to do counterfactual trends—is that if you look at the 1950s in the postwar era, men's wages are growing by much, much more than women's in this period. It's not exactly clear what that's about. Is it about increasing the labor supply of women to certain jobs and occupational segregation? Changes in the selection of women? There are well-documented increases in part-time work. We know that very different types of workers were coming into labor markets in that period. It could also be about increasing sex discrimination in labor markets. So all of those things were probably going on in the 1950s. If things would have continued as they had, we would have expected that the

gender gap in the 1960s and 70s probably would have grown instead of stabilized. So we think the stabilization of the gender gap is also very consistent with this anti-discrimination legislation really raising women's wages in a period when, all else equal, they wouldn't have fallen, but we would have expected them to fall further behind the growth in men's wages.

Steven: And that seems to be a very important point in general in discrimination studies—that one can think about the efficacy of legislation and successive various policies as holding back the tide, as it were, when they would have made things worse. It's very nice that it shows up in the analysis so cleanly. What conclusions might we draw about the current gender wage gap? One possibility is to say the legislation was successful, and so the residual mechanisms are going to require different remedies. The other is to ask whether the mechanism has been fully and completely identified. Of course, the truth is somewhere in the middle. What's your thinking?

Martha: I think both are true. One of the other questions we got a lot when presenting this paper—for example, Francine Blau's dissertation is really on occupational segregation. If you look at the work of a lot of influential labor economists writing in the 1970s and 1980s, they're really talking about how men and women are in very, very different types of jobs, and that this is a huge component of the gender wage gap. So we went back and coded up this occupational wage survey done in the early 1960s to look at whether, given very well-defined jobs collected as part of the survey, there was actually any within-job discrimination to remedy, or if all of the discrimination just happened across jobs. And in fact, we find that about half of the gender gap was within the exact same job as defined by the Department of Labor. Now, by the time Blau is writing her dissertation in the 70s, that's no longer the case. Part of the way the Equal Pay Act worked is that it encouraged compliance—women's wages were being raised in some jobs. You also see newspaper articles where firms were talking about how, while they will comply with the letter of the law in the short term, they plan to downgrade women's job classifications in the longer term. They literally said this in newspapers in the 1960s, so you don't actually have to figure out what they were thinking. By the time we get to the 1970s, in fact, a lot of the within-job pay gaps have really disappeared, and then the focus starts to be on occupational segregation and other things.

Steven: Much of the motivation of the War on Poverty was, trivially, the idea that poverty could be eradicated. But once you articulate the objective that way, it really becomes an intergenerational question—the extent to which policies that help the disadvantaged will create opportunities for their children, making it unlikely those children will be poor. You've done very important work on the long-run effects of Head Start, which is a clear example of a War on Poverty program whose intent was to break dynamic patterns. Could you describe your

work?

Martha: It goes back to my fascination with the 1960s. I remember I started working on this program when I was looking into some of the War on Poverty–era programs around women and contraception, and I kept stumbling across pronouncements on the right and the left across the political spectrum that the War on Poverty had decisively failed–that nothing was successful of the era. There were tons of prominent economists writing about this in the early 1970s. I was really intrigued by this because the standard of evidence today is so much higher than what was used in the 1960s to do evaluation. Social science was really in its infancy. A lot of what I spend my time thinking about is how we do better inference about the effects of public policies and programs. So I thought, well, how did they know that? Did they have the data? What was the research design? What types of evidence did they have at the time? A lot of this just kind of turned on the time series, which I find very unsatisfying. And I thought, well, my view of what we should use as a standard of evidence for whether or not these programs succeeded or failed and what was the view and maybe the politics of the 1960s and 70s are really different. That means maybe we should start digging into that a little bit. One of the first programs I started working on was Head Start. This program is one of the most widespread and most popular of the War on Poverty programs. Despite evidence that it didn't work, even dating to the 1970s, it has continued in the United States. I'm not saying that evidence was good evidence; I'm just saying that was the evidence of the day. The genesis of the program is really interesting. The idea is that public education should be widely accessible and can allow poor kids to escape their circumstances. But one of the things they noticed right away is that disadvantaged kids were starting school behind in math, behind in their alphabet, and often couldn't read as well as their more advantaged peers. So policymakers and educators started thinking about giving these kids from lower-income families a head start. That's how the program got its name. Another way they referred to it is "prep school for poor kids," which ended up in the title of our paper. The program rolls out all over the United States. There are two ways you could build a program this way–you could build an extremely intensive, well-resourced, but small program, or you could invest in setting up a program in every congressional district all over the United States, which was the path the Johnson administration chose. They went for broad coverage rather than program intensity or high funds per kid, especially when you compare it to programs like Abecedarian or Perry Preschool–Head Start has a lot fewer resources. Immediately, a lot of the educators who thought this was a great idea were put on the defensive. They thought this isn't high quality–they're not always hiring teachers with college or master's degrees, there are no standardized curricula, they're relying on a lot of parent volunteers, social workers, people without degrees in teaching to implement this program. So there was immediate backlash and calls to evaluate the program. There's a really early evaluation, the

Westinghouse Learning Study, which shows that the Head Start kids do much worse than a comparison group. But the issue turns on how they constructed the comparison group. It wasn't a randomized control trial; they went back after the fact and tried to identify other kids in communities similar to the Head Start kids and then made the comparison. Social scientists, of course, went back and found that the kids chosen for the comparison group tended to be a lot more advantaged in a number of dimensions than the Head Start kids, which is why the evaluation showed the Head Start kids were doing worse. That didn't kill the program. It kept going and even expanded. Then there's a quasi-experimental literature in the 80s and 1990s doing things like comparing siblings in the same family, one of whom went to Head Start and one of whom didn't. Those studies show that Head Start seemed to have benefited those kids in a number of dimensions. But, of course, there are concerns that maybe the sibling that went to Head Start did so because the parents were doing better, or they moved to a different, more resourced community, or a variety of other reasons why that would have been the case in the absence of the program. Fast-forwarding, we finally have the Head Start Impact Study in 2002, where they ran a randomized control trial. Parents first sign up to be in the experiment, then are enrolled into either a treatment condition where their kids get into Head Start or a control condition where they don't. This seemed like a great idea. But almost immediately those findings were critiqued. They find some evidence that Head Start benefits kids early on, but by the end of first grade, there's this fade-out in test scores, which sounds like a bad thing. But one of the things researchers have pointed out is that the parents engaged enough with their kids' learning to sign up for the experiment, but were assigned to the control condition really hustled and got their kids into other very high-quality preschool programs within a year or so, such that the difference between the treatment and control conditions in terms of access to good preschools is very, very small. The study is really not powered to detect an effect that small. So you have a fairly imprecise null; a lot of contamination of the control group, if you like. All that said, it's easy to say, well, what does Head Start really do if anything? This is the context where we decided to work on the program in the 1960s. The 1960s provide a much cleaner comparison. There weren't a lot of other preschool programs at the time, especially not for poor kids, and there wasn't even kindergarten in a lot of places. So kids showed up at first grade—that was it. Kids who were disadvantaged probably wouldn't have gone to kindergarten or preschool. So this is a comparison of Head Start versus the counterfactual of not being enrolled in any preschool program. The other advantage of studying this era is we can really sidestep the issue of what we actually want Head Start to affect. I don't think anyone thinking about anti-poverty programs as the ultimate goal really wants to raise test scores by some amount. The ultimate goal is to empower kids to escape their circumstances of poverty. And because we can look at kids' adult outcomes—those who would have had access to Head Start in childhood—we can think more broadly about the program's

effects on opportunities, life chances, and the cycle of poverty. There are a couple of challenges with this paper that we were able to overcome because of developments that happened right around the time I started thinking about the question. One is that the Census Bureau made it possible to link together modern census data through the entire 2000 Census and ACS data with information from the Social Security Administration, which gave us information on exactly when every child was born—allowing us to document their school entry age—and also the county in which they were born, which we can then link to whether or not that county had a Head Start program. So we can connect outcomes such as wages, occupation, whether you're receiving public program support today, the quality of the neighborhood you live in, your health or early mortality, whether you're incarcerated, to whether or not you had been in an area that would have had Head Start when you were ages three to five. The idea we were relying on is that the timing of when these programs started in different places was not completely random, but it's not really associated with a lot of other things going on. The fact that your county gets it in one year and another county gets it in another year; we're leveraging some of that variation. But we're also using the fact that we know exactly how old the kids are. Our comparison group in that paper is the kids that were just a little bit too old to have benefited from the program. They show up for first grade around the time Head Start is starting, which means they're too old to have participated. But kids just a few years younger—just because that's when the program ended up starting—would have showed up to first grade after the program had been in effect a few years and presumably could have benefited. Then we rolled us into a statistical model to account for a whole lot of other things. It was pretty exciting when we first saw the results. We spent a long time trying to kill them because we were worried they weren't right, because they were so positive. But overall, we see really large effects on children's education in a number of dimensions, including things like college and professional degrees, which hasn't been documented before. And we also see large reductions in adult poverty, particularly for women. Again, this is one of the original aims of the Johnson-era programs, but it's taken almost sixty years to document the effects.

Steven: That's also a beautiful example of the importance of economic history in elucidating contemporary questions and why economic historians are at the forefront of studies of inequality. One thing that comes to mind when you talk is this ex-post concern about test scores as ends in themselves in the evaluation of programs. That showed up in the literature on whether increased spending on students was efficacious. If memory serves me correctly, the original studies looking at the effects on test scores didn't find much, but then Card and Krueger re-examined the effects on labor market outcomes and found something. So I think you brought up a very important point in terms of actually identifying the objects of policy interest as opposed to intermediate variables whose relevance is because you have the data on

them as opposed to speaking to the substantive questions. I also wonder whether or not this obsession with the fade-out effect of test scores may have missed something. In other words, if the fade-out for a program at age four occurs at age sixteen, that means that person had twelve years of elevated whatever-it-is that produces the test scores. If that temporary boost were, for example, to facilitate a more rapid accumulation of vocabulary, then that's a permanent effect. So I've always found it a little puzzling that we don't have a clear vision of developmental trajectories in which the measures we have, like test scores, are really just noisy measurements about the things we care about—which is capabilities, and how they're getting exercised.

Martha: The people who write tests hope that they're writing tests that are associated with all these other things we actually really care about. The difficulty is knowing whether or not they are. When you see these fade-out effects, the worry is that they're not actually capturing the relevant things. Critics of this literature have told me that this is just another example of pop-up effects—you have effects that are there for a little bit and then they go away and pop back up in adulthood—and is what you're measuring really the effects of the program, or is this some spurious association? It's really hard to answer that. Historical studies are great in some ways, but we're really limited where you want to push us on the mechanisms. I don't have an RCT from the 1960s that's going to allow me to document every change we'd like to see. I'll also say something else I think is really hard, which is that our drive to get evidence-based policy often leads us to evaluate things in the very short term. You pass the policy, you have funding to evaluate it, you need to document changes showing up really, really soon. But the way that we think human capital works is that it accumulates over a lifetime; you have little changes right away in health and nutrition, in education and cognitive abilities, and these things continue to generate more gains over time. Just measuring them in the first few years, even if there are substantial effects, may be a dramatic underestimate of these lifetime gains.

Steven: Absolutely. It's almost a standard time series problem—you can estimate changes between three and seven, but the ability to extrapolate across the trajectory has a host of auxiliary assumptions that simply aren't going to be credible. So it speaks to the importance of getting the type of data that you're generating. Let me turn to your work on contraceptive freedom, contraceptive access, and women's socioeconomic outcomes. You're very much the leading scholar in this area, so I was hoping you could give a tour of the main findings of your research.

Martha: When I was in graduate school, I was just enamored by a lot of Gary Becker's work. I thought it was brilliant. I loved thinking about human capital and applying the tools of economics to topics like

childbearing. But I remember reading through one part of Becker's treatise on the family where he says the birth control pill really hasn't been that important at all in changing anything—he's making the case that it's really not what's driving fertility. And I remember thinking, I don't think that's right. Most of what Gary Becker wrote I thought was very, very sensible, but that was one of those topics I thought couldn't be completely right, and I started thinking about how we could do more to get evidence on this. Just to rewind: there's a whole camp of economists going back a long time who have been on record emphasizing that the demand side of the market—income and prices—is doing most of the work in terms of driving childbearing over the longer term. The idea is that as incomes rise, couples tend to substitute away from large families toward smaller families and make more investments in their children, something Gary Becker called child quality. At the same time, fertility rates are also falling because the time costs of raising children have risen very quickly, because women are now paid much more in labor markets to do other things—so the opportunity cost of child-rearing has risen a great deal. On the other hand, I picked up this volume by Charlie Westoff and some others about the contraceptive revolution, and demographers were arguing that modern contraception had changed the world and played a central role in declining childbearing, none of which was really in the Becker model. I was puzzling about this, and as a woman, I had a hunch that there was something to the contraception story that wasn't being told in Gary Becker's treatise. The implications of this are hugely important because developed countries have all gone through this demographic transition from high fertility and high mortality to low fertility and low mortality, and as countries think about how to accelerate their transitions to industrialization and more developed economies, thinking about how to transition demographically is also at the forefront. Why do countries transition? What are the policy implications? And of course, the implications for gender inequality. One of the things I picked up in graduate school was Claudia Goldin and Larry Katz's important work on the birth control pill, called "The Power of the Pill." It was exciting to think about and then start to think about what I'll call the role of the supply of contraception in shaping the modern economy in the United States and women's experience. And of course, it's really hard because the birth control pill became available at the peak of the baby boom, and fertility rates were probably going to go down anyway, so the empirical challenge of trying to unpack what's the causal effect of the pill versus what's just the natural evolution or mean reversion we would have observed in the absence of this technology is pretty tough. So my job market paper was my first attempt to quantify the importance of modern contraceptive technology, borrowing heavily from Goldin and Katz's research design. What they were using is this: the legal age of majority in the United States in the 1960s was 21. In large part because of the Vietnam War—men were being drafted at age 18 to fight, but couldn't vote about whether they'd be drafted until they were 21—states had started to change the age of legal adulthood from 21 to 18.

What that did, as Goldin and Katz point out, is it allowed younger women age 18 or 19 to consent for medical care as legal adults, whereas in states that hadn't had these legal changes, they had to wait until they were 21. I call this early legal access to the pill; the idea is whether speeding up when women could get this technology, which would allow them to delay marriage and delay childbearing, affected the ways they invested in their careers. Thinking about what was going on in the period, you have a lot of the marriage market clearing around age 22 or 23, and the idea is if you missed that boat you might not be married, you might be one of those "old maids" that you didn't want to be—which wasn't just something you didn't want socially but was also economically a really bad outcome because you needed to rely on your husband's income. So if you could delay childbearing just a little bit, stay in school, pursue more demanding majors or a different type of occupation—did the pill really facilitate that? And if so, did it not just shift decisions in the short term but reshape the entire equilibrium of when people started to get married, how long women stayed in school, and alter their career trajectories? That's one of the things I went after in my job market paper—not just whether this affects childbearing or marriage for a year or two, but whether it kind of reshapes the way women are making these investments. One of the things I documented is that this change in access to the pill at earlier ages is really associated with changes in the shape of women's labor force participation profiles over their lifetime. Rather than the old norm—getting married, having a child in your 20s, withdrawing from the labor market for that full period, then re-entering when your children start school or when they leave home—women started to both work and have children more often at the same time, re-entering after each child and timing them in accordance with their career aspirations. I go on in a follow-on piece with Brad Hershbein and Amalia Miller to use the NLS Young Women survey to see if we could observe changes in women's investments in occupational training and other types of human capital decisions associated with that. So this first suite of papers is really about how younger women changed their plans about family and career with the introduction of modern contraception. But I didn't stop there—there are three different papers about this. There are two more: one is about married women, and another is about disadvantaged women. The idea for married women is that if you've already gotten married and made your career investments, we don't think the birth control pill showing up is now going to make you into a lawyer or a doctor—you've kind of already made those decisions. For them, we wanted to look at whether it changed the number of children they had. So, for that, I looked at the Griswold decision, which is a Supreme Court decision in 1965 that struck down Connecticut's ban—unique in its ban on the use of contraceptives for married women. It turns out there were over 20 other states that also had bans on the sales of contraception, which would have made it more difficult for women to access the birth control pill. Even though the Griswold decision didn't make those laws unconstitutional, states rapidly removed them in the aftermath of the

decision. And you see changes in women's childbearing really corresponding to that—large increases in the use of modern contraceptives in states where the sale of contraception had previously been banned after the Griswold decision, and then large declines in birth rates. In short, it's a way to get at how much of the decrease in childbearing coming off the baby boom is really associated with this contraceptive technology versus how much is just the mean reversion we would have seen in the absence. The other piece I was really interested in is how family planning programs during the War on Poverty affected the birth rates of lower-income women. Goldin and Katz's piece was really focused on married women and college women because a lot of the way these women were getting access to the birth control pill was through their college-level clinics—already more affluent and educated groups. But what about everyone else? It turns out a War on Poverty program introduced in 1965 was to fund family planning programs that rolled out to clinics all over the United States. Started under the War on Poverty, it eventually became a national program in what's now known as Title X of the Public Health Service Act, which continues to fund clinics today like Planned Parenthood. The idea I had is that even if legal access wasn't an issue, financial access would have been a really big issue for lower-income women and would have really limited their ability to use modern contraceptive technology, even if it was something they very much wanted. One of the things we can do in that paper is document that these programs start at different times all over the United States, and then you can see that the start of one of these programs is associated with fairly large reductions in birth rates. The other thing we've gone back to do is look in the Census at how much better the children fare in terms of their parents' resources, the likelihood that their parents are together in two-parent households, the receipt of public funds. And how much those things vary by whether the child was born after a family planning program was begun versus before. One of the things we find is large increases in the incomes of the average child that is born, and reductions in the likelihood that the average child will live in a single-parent household or in a household that collects public assistance. This is hard to interpret because it's easy to say inadvertently that these programs increased the incomes of parents or reduced the likelihood of receiving benefits, but that's actually harder to say because of selection effects. What I mean by that is that you can have programs that reduce birth rates for certain groups of people that increase the average of some outcome for children but don't necessarily increase the income for any child that is actually born. So we go through and try to bound that for the paper too. And it’s something an editor made me edit out, and I actually really loved this: I called it the empowerment effect and the selection effect. The empowerment effect says if you can use contraceptive technology to delay childbearing just a couple of years—to meet the right person, maybe get into a long-term stable partnership, find a better job, increase your income—that's the empowerment effect. That is empowering parents to become parents when

they would like, and children are then born into better-resourced environments. The alternative is that children are simply missing from certain groups, which I call the selection effect. We think a good bit of what we're documenting based on some bounding exercises is really due to this empowerment effect, which I think is really important but has largely been ignored in a lot of the literature around family planning, and especially abortion. The bottom line of a lot of the research I've done is that modern contraception in the United States unleashed something of a demographic tsunami. We have rapidly trended toward below-replacement fertility rates, delays in childbearing into later ages, shifts to later marriage or cohabitation rather than marriage, and we've seen a rise of women through the educational and professional ranks. I don't see these as things we're going to reverse anytime soon.

Steven: Martha, thank you so much for an absolutely fascinating conversation. And let me thank the audience for listening.

Martha: Thank you, Steve.

Steven: One of the interesting phenomena associated with contemporary gender inequality is that on many measures of educational success, women are outpacing men. Just to give one example, an analysis from the Pew Research Center published last year found that 47% of women aged 25 to 34 have a bachelor's degree, in contrast to 37% of men in the same cohort. And yet despite this comparative academic success, women often do not have the same success in labor markets. Natasha Quadlin, professor of sociology at UCLA, has conducted ambitious experiments that examine how men and women's experiences in the labor market diverge. The results are fascinating and important, and she discusses them at length with my colleague Geoff Wodtke.

Geoff: So you've done a lot of research on gender differences in academic performance and their consequences. I wonder if we could start with you telling us a little bit about these differences—how large they are, when they first emerge, do they differ by subject?

Natasha: This is a great question and a really good place to start our discussion. The short answer is that we see notable gender differences in college academic performance. College enrollment itself is quite complicated; students are in different majors, which are very much gender-segregated, and they're in different types of institutions where grades might not be comparable. Rates of non-completion in college are also quite high. But most angles of this comparison tend to advantage women. Women have higher grades in college, they attempt and complete more credits, they're more likely to persist in college, and they're more likely to complete bachelor's and advanced degrees. Some of this is explained by the fact that men tend to enroll in “harder subjects” like math and science, or STEM majors that remain heavily male-dominated and are also associated with higher pay and

prestige. This is one of the reasons I wanted to study the effects of grades within STEM majors, which we'll get to in a second. But there's even research showing that among students who persist in math and science classes, women have comparable grades or even a slight advantage in some studies. So, really on every metric, in any way you slice college academic performance, women tend to have an advantage. Research also provides many explanations for why this happens, and the answer generally seems to be that from the time children are first formally assessed in school, girls are rated as having stronger skills, and these gender gaps tend to persist and often grow over time. Even when girls and boys are in kindergarten, girls are rated as having more self-discipline and being more attentive. This continues into middle school and high school, where girls' grades are much higher, and girls are much more involved in school. This gender gap in high school academic performance tends to explain much of the gender gap in college academic performance, as well as the gender gaps in college enrollment and completion. Importantly, this really doesn't jive with women's broader status in society, right? Women might be seen as very good at school, but then less well-suited for the labor market. And I don't know of any research that assesses the general public's assumptions about the gender gap in college completion and where it currently stands. I actually think that most people might assume that men still have an advantage in college completion because men are generally seen as higher status. So there are these very well-established patterns of girls and women excelling academically, but whether people really know about this is more of an open question.

Geoff: Do good grades or academic performance in college matter the same for everybody, or are there reasons to suspect that employers might perceive the achievements of different types of students in different ways?

Natasha: Gender is the key criterion that might determine what exactly good grades might mean to an employer, and that's what I focus on in this study. We might expect that employers might be more skeptical or have more questions about women who get good grades, and these explanations come from social psychology and from well-established traditions. The larger tradition is called the stereotype content model, specifically, a theory that tells us about what happens when we encounter men and women who we perceive as competent or smart, like a job applicant who has good grades. The theory tells us that men can be readily perceived as both competent and likeable simultaneously. When we perceive a man as competent or smart, this doesn't impede our ability to think that we like and admire this person—we might readily see a competent man as a potential friend, co-worker, or leader. But the way that we evaluate women is quite different. When women present as competent, a common response is to question their likeability or to have some kind of trade-off between competence and likeability. We may see competent women as more socially threatening than competent men—they can impede on our social position—and so we might put our guards

up and not see women as potential friends, co-workers, or leaders to the same extent. I also want to say this is true of both men and women as evaluators. A common question I often get is whether this is how women evaluate women, or is it just men who aren't receptive toward competent women. The answer that emerges in so many studies is that it's both, because we all live in this world governed by strong gendered norms and expectations, and so it's not just men who carry these beliefs. The other thing I want to mention is that this isn't just a gender issue—it's also a race issue, and therefore an intersectional issue. We have strong assumptions about white boys and girls' performance and about Black boys and girls' performance and Asian boys and girls' performance, et cetera. The stereotypes we have about groups are widespread, and gender and race are extremely important in shaping expectations for how people perform and how they behave. These expectations can have consequences that are sometimes even stronger than a person's actual performance.

Geoff: Could you tell us a little more about the details of the studies where you've investigated how employers respond differently to academic performance among men compared to women? How exactly did you go about figuring out whether grades matter differently for these two groups?

Natasha: I wanted to study this in a way that would tell us not only how employers evaluate job candidates but also why they evaluate them the way they do. To do that, we have to combine a couple of different methods that give us different pieces of the puzzle. First, I conducted a study known as an audit study or correspondence study. These studies are common in sociology and economics. The general idea is that you apply to a large number of jobs while posing as different applicants, and you're able to track which applicants do and don't get called back for an interview. Using this process, I applied for about 2,100 jobs. When creating these fictional applicants, I was strategically manipulating three things. First is their gender—half are men, half are women. Second is their grades—anywhere from a 2.5 to a 4.0. Third is their major—and majors are heavily sex-segregated, so the assumptions we make about students' skills and personality could vary not only by their own gender but also by the gender typing of their major and the interaction of those two. In this study, students were either business majors, English majors, or math majors. Business is more gender-neutral, English is a humanities field that's heavily female-dominated, and math is a STEM field that's heavily male-dominated. Between these characteristics and being able to observe how applicants with all different combinations are being evaluated in the labor market, this potentially tells us a lot about how entry-level job applicants might be evaluated. For the second part, I also conducted a survey experiment, where I presented people with different stimuli—in this case, these resumes. I showed each person a different version of one of these entry-level resumes and asked them the same questions so that, depending on the resume the person saw, I could see

how their answers varied. This is something more researchers are starting to do when they conduct audit studies because an audit study doesn't tell us all that much about why applicants are or aren't being called back—all you get is a callback or not. A survey experiment is how we try to get around that. I recruited a sample of about 300 hiring managers, presented them with one of the same resumes I used in the audit study, and essentially said, what are your impressions? How do you rate this person in a hiring scenario, and why? To what extent do you perceive them as competent, as likeable, as dependable, do you think they'd have an outgoing personality with co-workers and clients? I asked a lot of targeted questions about personality traits, and I also had respondents write free-response answers about how they perceived the applicant. So you get quite a variety of data between these two different studies.

Geoff: Which of those applications was most likely to get employers' attention, and which were least? What did your survey respondents tell you about what they were inferring about these hypothetical applicants based on their grades, major, and gender?

Natasha: I found a lot of key differences, and some of these were quite striking. Let's start with overall gender differences before talking about how these differed by major. Overall, I found that grades for men really didn't matter that much—which was very much what conventional wisdom suggested. When I put all these majors together, I didn't find differences in callbacks between low-achieving, moderate-achieving, and high-achieving men. For women, there was a very different pattern overall. Initially, women's callbacks went up—if they had moderate grades, they tended to get called back more than if they had low grades. But high grades didn't provide an additional advantage. The callback rate between moderate and high-achieving women actually goes down. Importantly, this also varies by field of study, with a particularly consequential pattern among math majors. If we just look at math majors specifically—a heavily male-dominated STEM field—we see that men actually do benefit from having high grades. As men have higher grades in math, their chances of getting a callback go up and up. But again, this is not the case for women. High achievers are called back very infrequently. The result is that high-achieving men in math are called back significantly more often than women, at a rate of about three to one. The next question is why does this happen, and the survey experiment data point to very different preferences for men and women job applicants in terms of personality traits. Employers tended to say they wanted men who were competent and committed, which is a perspective that's good for men with high GPAs. Men may not always benefit from having good grades, but if anything, it doesn't hurt them—employers are willing to give the competent and committed man the benefit of the doubt. For women applicants, however, the trait most strongly associated with callbacks was likeability—not competence or hard work or commitment, which all theoretically matter for job performance, but likeability. This helps explain why employers may

prefer moderate-achieving women to high-achieving women. These women with more middling grades are potentially likeable and seem like they'd be good coworkers, but employers were often skeptical of high-achieving women's personalities and their ability to work well with other people. Ultimately, this is how high achievement may be converted into a liability within a professional setting and hurt women's prospects.

Geoff: What are the implications of these patterns for downstream outcomes like gender gaps in earnings, or maybe even outcomes that come during college when men and women are deciding which majors to specialize in?

Natasha: There are a couple of things I would note about the gender earnings gap specifically. This provides evidence of barriers for high-achieving women in the labor market, which others have written about extensively, especially the important research that points to women's difficulty obtaining leadership positions not only in the workplace but in society. But this isn't a bias that's necessarily been linked to grades before. This type of attitude is likely to affect women at various points in their professional trajectory, whether it's bias at the point of hiring, which I focus on here. or a similar attitude that's emerging in women's performance appraisals or considerations for leadership. Second, to talk about the wage gap specifically, this is a part of the paper I don't usually discuss as often but it's relevant. I essentially consider whether employers are discriminating against high-achieving women because they perceive them as overqualified. This is one possible explanation for the patterns I find-and you could twist this into a backhanded compliment in some ways, which is often how gender bias works. I borrowed a method from one of David Deming's papers to essentially assess how women's callbacks vary for jobs at different points in the pay scale. The logic is that if high-achieving women are perceived as overqualified or unattainable, they might not get called back for the worst jobs or the middling-pay jobs, but maybe they'll get called back for the best jobs, where they might be seen as attainable. I test this explanation and don't find support for it. If anything, high-achieving women are called back even less often for the highest-paying jobs, which I see as a really problematic sign of double disadvantage. I've published research related to this, especially a paper with Tom Vignhuvlin at the University of Minnesota, where we look at earnings data among millennials and find that there are large gender wage gaps that emerge at the very outset of employment for STEM workers specifically, but not for non-STEM workers. For non-STEM workers, these wage gaps are much smaller and take a while to grow. So this is another related pattern with implications for the gender wage gap, especially among STEM workers, where the biggest gaps are emerging.

Geoff: What are some of the policy implications of your findings? What would you recommend to help mitigate gender disparities in

specialization in college and encourage talented women to pursue those majors and occupations when they finish school?

Natasha: There's a lot that we can do and a lot that we've already been doing. The implications for STEM majors and occupations are particularly important. Gaps in college majors have persisted for many decades, and women getting feedback that their performance is not going to be rewarded in the same way as men's is not feedback that's going to help. There's a lot we can do within education—encourage women to choose STEM pathways and persist in them, with efforts at all levels of education from elementary school through post-secondary. Research shows it's very important to help girls develop a math identity and maintain a high sense of self. But obviously, efforts within education are not enough. It's not just on higher education to fix this problem. Employers need to continue to develop policies and programs that will recruit and support women in STEM fields. It's not nearly enough to work only on the side of education—it's an employment problem as well. But I also think universities can play a role here, and I've been thinking about this a lot lately. Universities really only have a hand in the entry-level labor market, but they're also an important source of labor within local labor markets. I think there's an argument to be made that universities can take a much more active role in ensuring equity and inclusion in the employment of their graduates—whether through closer employment partnerships or policies that can be put in place to ensure equity in on-campus recruiting. I'm personally hopeful about these and excited about how universities can maybe be part of this effort to reduce inequality.

Geoff: All right, well thank you so much for a really interesting discussion.

Natasha: Thank you, Geoff. Thanks for having me.

Steven: The Inequality Podcast is a production of the Stone Center for Research on Wealth Inequality and Mobility at the University of Chicago. I want to end with thanks to the people who really make it happen. First, I want to thank our producer and engineer Shane McKeon. Second, I'd like to thank our assistant director Gerardo Espinal Franco for the production oversight and doing everything that is required to bring the podcast to fruition. And finally, I'd like to thank our executive director Grace Kolavo for her support, not just for the podcast, but for every activity at the Stone Center. You may get in touch with us at stonecenter@uchicago.edu. Thank you so much for listening.

This transcript has been prepared with the assistance of AI.