THE SUPPLY OF GENDER STEREOTYPES AND DISCRIMINATORY BELIEFS

by

Edward L. Glaeser[1]

Harvard University and NBER

and

Yueran Ma

Harvard University

December 5, 2012, Preliminary Draft

Abstract

An overwhelming majority of men and women born early in the 20th century believed that was better for women not to work; now a majority believes that work is appropriate for both genders. It seems reasonable to think that beliefs about women's ability levels have changed as well. This paper presents a series of model examining the social formation of beliefs about women's roles and ability levels. Our first model follows Betty Friedan and examines the formation of beliefs by consumer good providers. We conclude that they only have the incentives to supply error, when their products complement women's time in the household, and when production is highly oligopolistic. A second model focuses on formation of beliefs by in-group workforce members eager to discredit possible competitors. Finally, we turn to belief formation in the household where husbands and parents perpetuate stereotypes in order to encourage women to spend more time in the household and to have more children. If children overestimate their parents' level of altruism, then parental misinformation will be more effective and prevalent. Theory predicts that the 20th century postponement in female child-bearing may strongly reduced the supply of gender-related misinformation, because delay provides hard data which reduces the power and prevalence of false beliefs.

[1] Glaeser thanks the Taubman Center for State and Local Government. This paper was written for a conference celebrating Claudia Goldin and it was deeply inspired by her work.

I. Introduction

What causes gender-related beliefs to emerge and shift over time? According to the General Social Survey waves of 2003, 2004 and 2007, 47 percent of women born before 1946 (and 59 percent of men) agree or strongly agree with the statement "It is much better for everyone involved if the man is the achiever outside the home and the woman takes care of the home and family." Only 29 percent of women born after 1945 share that view. A full 50 percent of female respondents (from all cohorts) agreed with that statement in America's West South Central Region, while only 26 percent of New Englanders shared the view.

We have less wide-ranging direct evidence on discriminatory beliefs about women's ability in the workforce than we do about women's "proper" role in the home, perhaps because surveyors may not have trusted respondents to answer truthfully. Nonetheless, the evidence that does exist also suggests dramatic transformations about beliefs about women's capacity during the late 20th century. In 1953, Gallup asked "If you were taking a new job and had your choice of a boss, would you prefer to work for a man or for a woman?" In the 1953, 57 percent of women and 79 percent of men expressed a preference for a male boss, as opposed to only eight percent of women and two percent of men who expressed a preference for a female boss. By 1987, the share of female and male respondents expressing a preference for a male boss had dropped to 37 and 29 percent respectively, with men now preferring a female boss (Simon and Landis, 1989).

Moreover, an abundance of personal histories, ethnographic work and field-specific statistical work suggests that men, and often women as well, have believed that women are less capable in many-workplace relevant tasks (e.g. Lerner, 1987). The literature on women and perceived math ability is voluminous, and suggests that men and women often both believe that women are less able in mathematics (see Gunderson et al. 2012 for a review). The women who were pioneers in corporations have often described a common male presumption that their talents were limited. The pages of literature are replete with stories of women who experienced low expectations in their own homes, from both spouses and parents. Ibsen's "A Doll's House" provides a classic example.

This paper does not attempt to add new measurement of discriminatory beliefs, nor does it attempt to quantify the effect that such beliefs may have had on women's labor force successes or family outcomes. Our goal here is to take the survey, ethnographic and literary sources at face value and assume that patriarchal and discriminatory beliefs have existed and that they are important enough to better understand. Major thinkers from Aristotle to Freud have often depicted women as severely lacking in vital decision-making areas. We furthermore assume that these gender-related stereotypes cannot be understood as a Bayesian response to facts. The surveys discussed above are taken in the same year, by respondents who observe the same labor markets, and yet respondents born before and after 1945 have markedly different opinions about the advantages of work for women. At a particularly general level, there is every reason to believe that male attitudes towards women's innate ability have shifted dramatically since the

Victorian past, and it is hard to see how women's "innate" characteristics could have experienced any such dramatic shift.

Instead of basing beliefs on unvarnished facts, we assume that beliefs reflect persuasion rather than reality, as in Glaeser (2005). While this work builds on that earlier paper, there are two major conceptual differences between this paper and that essay, which focused on the formation of group-level, ethnic hatred. First, hostile racial beliefs often focus on the threat caused by the out-group, or the "untrustworthiness" of that group in the language of sociology. Conversely, stereotypes about women rarely suggest a threat, but rather limited competence in the workforce and remarkable productivity within the home. Second, in the case of racial hatred, it was easy to find entrepreneurs of error among the politicians who spread misperceptions as part of their jockeying for power.

Indeed, the primary challenge facing the analysis of the supply of gender stereotypes is to find the primary promoters of gender stereotypes, for different sources have emphasized different culprits. Friedan (1963) looked to magazine publishers and companies selling products for the home that were presumably complements with women's time in the household. Other sources have suggested that parents or teachers play a larger role in perpetuating gender stereotypes (Gunderson et al. 2012) . Still others, like Ibsen, have pointed to spouses who wanted a subservient or home-oriented spouse. Finally, others have credited jealous co-workers with spreading tales of female incompetence.

We model all three of these scenarios, hoping that the predictions of these models will eventually face the data and help to determine the relevant importance of different sources of gender-related beliefs. In all three cases, we will discuss the generation of beliefs about women's ability in the workforce, but for the first two types of models, work and home-related returns may be non-pecuniary, and misperceptions may not relate to enhanced expectations about the joys of home life, not stereotypes about female incompetence. When beliefs are created by co-workers, then they must direct pertain to women's competence to be effective.

Our first model follows Friedan (1963) and emphasizes the generation of beliefs by sellers of household goods. We assume that these companies have access to a technology that generates apparent examples of working women experiencing negative outcomes. This assumption is meant to capture Friedan's description of magazine articles, generated by male publishers perhaps eager to please their advertisers, which depict happy housewives and miserable career women. We assume that women discount many of these stories, but as long as the marginal story is given some credence, then they will have some effect on the amount of time women spend at home or work.

The model suggests that gender stereotypes will be strongly dependent upon the market structure in consumer goods, the effectiveness of communication technology and the naïveté of women about third party stories. In order for Friedan's model to work, the costs of persuasion must be low, the household goods market must be extremely oligopolistic and household products must complement, not substitute, women's time in the home. Since manufacturing lifestyle norms is an industry-wide public good, a highly competitive industry will not do much of it. Arguably, the late 1950s combined large market power held by a few large home product producers, more effective means of communication and limited skepticism about that communication, yet those conditions are unlikely to have held throughout the long history of patriarchal beliefs. Moreover, it is surprising to see more counter-persuasion during that period provided by producers of technologies, like the washing machine, the dishwasher, the microwave and the vacuum cleaner, that substitute for women's time in the home.

Moreover, it is hard to see why fomenting gender related stereotypes where the cheapest means of persuading married women to buy new cookbooks or cooking appliances. It would seem more likely that the providers of home-related products, whether complement or substitute for women's time, will broadcast the simpler messages that a nice home can be nice. Indeed, such banal messages often seem like the stuff of home product advertisements, and these messages imply little about whether women are capable of finding fulfillment in the workforce.

The second model concerns the production of stereotypes in the workplaces by ordinary male co-workers who seek to avoid labor force workplace competition. The model assumes that men believe at some point in time, a colleague will be promoted and have the opportunity to replace him with a women. Ordinary workers play the role of advertisers of the first model, and they can, at a cost, spread stories detailing female incompetence in their specific occupation. We assume that the stories are true, but unrepresentative. The primary behavioral quirk is that individuals who hear stories remember them but occasionally forget their highly motivated source. It is also possible that listeners underestimate their co-workers' willingness to strategically deceive.

Stories are again spread only when they are low cost, and the cost of an ego-bolstering tale of female incompetence may be quite low. Moreover, the possibility of targeted persuasion is quite high. While the purveyors of tools for better being a better hostess may want to induce a wide swath of society's women to stay in the home, but workers really only want to influence a possible future boss.

Spreading false beliefs will be more common when women really are a potential threat, and this means that we can make sense of the rise and female of discrimination in certain jobs that is discussed by Goldin (2006b). During the early 20th century, the threat of a female competitor was small and this meant that men spent little effort on persuading prospective bosses not to hire women. During the middle years of the 20th century, the threat became more obvious and men began to persuade more assiduously. At the end of the 20th century, there were enough examples of real women working that misinformation had much less effect.

This dynamic also has implications about the relative persistence of gender-related quotas that limit the number of women on the job and glass-ceilings that prevent women from rising above a certain level. Gender related quotas should be unstable, if they are sustained with incorrect beliefs, because the few women hired for the job end up providing information that counteracts false beliefs. Glass ceilings, by contrast, provide no such evidence, allow false beliefs to persist and maintain the incentives to perpetuate such beliefs.

This model also sheds insight about the spread of discrimination in small and large groups. As group size increases, the incentive to spread misinformation diminishes, because the relevance of any one target of persuasion diminishes. Yet the total amount of persuasion may still increase, but the number of persuaders is also going up. This uncertainty implies that inaccurate beliefs may be more or less common in large social organizations, such as cities.

Section V then turns to persuasion in the family. The first model, which is inspired by Ibsen, examines husbands' incentives to persuade women that they are less competent. Husbands manipulate wife's beliefs by destroying evidence of the wife's ability. In this model, there are no irrationalities and average beliefs are correct, but some women still have inaccurately low estimates of their own ability (or they pleasure that they will get from working). Men want to spread inaccuracy if they benefit less from women working than women do, at least relative to women working at home. One possible reason for this difference is that women have more control over their own earnings from work.

Section VI looks at belief formation by parents within the family. Parents are altruistic towards their children but they have an independent desire to have more grandchildren. This desire creates an incentive for them to try to generate beliefs that lead to more child-bearing. The first model focuses on daughters' beliefs and daughters' education. If education increases the returns from working in the labor force relative to child-bearing, this will generate lower levels of women's education, even if women know their ability levels with certainty.

The under-provision of education effect gets more pronounced if parents, but not their daughters, have private information about their daughters' ability levels. In that case, parents of skilled daughters may have an incentive to try to imitate parents of less able children by giving them less education. If daughters have rational beliefs, this imitative will again cause more able women to think that they are merely average, but will not lead to any aggregate misperception about women's ability.

If daughters make the understandable error of overestimating their parents' altruism towards themselves, then the situation can become more extreme. Trusting their parents too much leads daughters to underestimate their parents' incentive to act strategically. This tendency will heighten the parents' incentive to behave in a strategic manner, by under-investing in education. Daughters may end up believing that they are in a separating equilibrium, when only the parents of the less able provide little schooling, while they are actually in a pooling equilibrium, where all parents provide little education to their daughters. .

In the second half of Section VI, we turn to maternal lifestyle choices and the impact of those choices on the beliefs of both sons and daughters. Again, we assume that parents have some private information, in this case about the relative returns to working at home or in a job. Mothers can signal, through their own lifestyle choices, the relative attractiveness of each lifestyle choice. Able mothers, who want more grandchildren, may avoid work in order to convince their sons and daughters that traditional lifestyles are best for women.   The impact of this signaling, and its prevalence, will again increase if children overestimate the extent of parental altruism and consequently underestimate the strategic role of parental choices.

Section VII turns to the timing of work and child-bearing. In this model, women have the choice of when to schedule a continuous term of home production for producing children.   One disadvantage of postponing child-bearing is that it leads to a shorter, continuous time of work, which limits human capital accumulation. A second disadvantage reflects potential from health risks from delaying child-bearing. In the model, the major advantage of postponing is that women can learn their ability levels if they work during an earlier period, which enables them to make better decisions about the tradeoff between parenting and work.

As long as the desire to eliminate breaks in work history isn't too strong, then a woman's decision to have children immediately or to wait depends on the state of medical technology, as discussed by Goldin (2006a). In the model, reduced risks from late child-bearing will delay child-bearing and lead to more information at that decision-making stage.   The model then embeds this choice in a framework that allows for belief formation. The critical implication is that investment in misinformation makes sense when women have kids early but not late. This fact implies that the shifts in the timing of women's child-bearing should have had a major effect on the supply of gender stereotypes. Section VIII concludes.

II. Discrimination and the Social Formation of Beliefs

We begin this preliminary section with a brief discussion of beliefs about women working and then turn to a more general discussion of the social formation of beliefs.

Beliefs about Women and Work

We have a great deal of information about women in the workforce, including the relative productivity of men and women in the household, the availability of market-provided household services, and perceived workplace discrimination against women (e.g. Goldin, 1990, Blau et al. 2011). We have less evidence on beliefs about female competence. Perhaps this dearth of information is understandable. After all, in the modern world we would hardly expect respondents to admit to gender-related biases. Nonetheless, the relative absence of polling data about female competence makes it difficult to fully document shifts in beliefs about women and their capacities.

There is however a great deal of more "anecdotal" evidence suggesting that women have often faced strong belief-related barriers to employment. Men have often held strong opinions that women were just not up to certain jobs. Often, these beliefs have crumbled in the face of reality, but certainly some of these beliefs persist.

In this subsection, we briefly review the polling data that is available about gender stereotypes from the General Social Survey and other sources. The General Social Survey, and other surveyors, has been asking questions about traditional gender roles since the early 1970s. Unfortunately, these gender-role related questions do not map clearly into any particular taste or belief. A patriarchal viewpoint can reflect a higher opinion of female productivity in the household sector, or a belief that employers discriminate unfairly against women.

Figure 1, for example, shows the average responses to the question "It is much better for everyone involved if the man is the achiever outside the home and the woman takes care of the home and family" by birth year for men and women separately. The graph shows a strong downward pattern for both men and women. For cohorts born at the start of the 20th century, almost all men and women thought that traditional gender roles were best. The share of respondents sharing that view declines to about 30 percent by 1950 and then levels off. There are some odd positive upticks in the responses to the question in the most recent cohorts, but this may reflect measurement error. Certainly, the basic pattern documents a profound change across cohorts born in the first half of the last century, and this pattern presents itself during every year in which the survey question was asked.

The second figure shows a similar response to the GSS question asking whether mothers' working outside the home is harmful or harmless for young children. Again, cohorts born at the start of the 20th century almost uniformly believed that children were hurt by women working outside the home. By 1960, almost half of respondents did not state this belief. Even though an overwhelming majority of respondents say that women working is just fine overall, a modest majority still say that working while children are young harms children.

There are far fewer questions that seem to directly capture assessments of female competence, and most that are relevant concern very particular tasks or occupations. The General Social Survey asks three highly specialized questions, in individual years only, that would seem to relate to female competence: the first (asked in 1974 and 1982) asked if men make better political leaders.

The cohort pattern, shown in Figure 3, is clear. About 50 percent of women born earlier in the 20th century think that men make better political leaders. By the latter decades of the century, this belief is down to 20 percent. We cannot generalize from political competence to competence in the workplace, but the effects are still quite striking.

A second question that may reflect on assessments of female abilities was asked in 1982. Respondents were queried as to whether female soldiers would help or hurt the military. Of course, it is possible to believe that women are just as able as men, but that their presence in the armed forces would lead to coordination problems and distraction. The sample is quite small for this variable, so we have grouped responses by decade in Figure 4, rather than by birth year. There is a downward pattern, although the most striking drop occurs at the very start of the 20th century and sample sizes are small during this time period. Overall, very few respondents seem to believe that women are harmful for the military, although surveyors may have got a different answer to this question if they asked it in 1940.

A third question that is potentially related to ability was asked in 1996. Men and women were both asked if women earn less than men because they work less hard. This question about female work effort shows a striking non-linearity (shown in Figure 5), where beliefs about greater male effort decline with year of birth during the first half of the twentieth century and then a rise after that date. We have no real explanation for this pattern, but it does suggest that cohort does have an impact on these beliefs.

The final two questions relate to women's outcomes in the workforce rather than beliefs about female competence per se. Instead, these beliefs presumably reflect assessments of gender stereotypes or other barriers to female labor force success. In Figure 6, we show responses to the question as to whether gender aids at work. Positive responses to this question decline with year of birth. Figure 7 however looks at the responses to a similar question: are jobs worse for women than for men. This figure shows little discernible time pattern.

The Social Formation of Beliefs

We now discuss the connection between our models and the broader economic literature on discrimination. The economics of discrimination began when Gary Becker (1957) presented a model of discrimination based on the preferences of employers, customers and fellow workers. The literature was extended by models that focused on a society-wide equilibrium that restricts the choices of a disadvantaged group (e.g. Krueger, 1963, Thurow, 1969, Akerlof, 1976). A third core model of discrimination, which owes much to Arrow (1973), emphasizes beliefs and statistical discrimination. .

Becker's approach to discrimination posits that some members of one group dislike working with or buying from members of another group. The Becker model does an admirable job of describing the reality of the mid-1950s, and provided a number of keen insights, like the negative impact on profits generated by an employer's discriminatory tastes. Lazear's model of culture and language provides a complementary communication-based explanation for some forms of discrimination in the labor market. Difference cultures, or ways of speaking, can make coordination difficult and lead to lower productivity. The costs of incorporating groups from different cultures into a single firm will lead to segregation, just like discriminatory tastes of employers or employees,

But even if whites had no innate dislike of blacks and men were willing to work with women, members of one group might still benefit if they were able to coordinate to expropriate the rights of another group. The South's Jim Crow system was not merely the decentralized preferences or beliefs of ordinary people. It was socially and legally organized, and seems in many contexts to have generated transfers from blacks to whites. Those transfers were perhaps most obvious in the case of segregated schools, which allowed tax dollars to be spent far more heavily on white children, rather than black children.

Krueger (1963) and Thurow (1969) both argue that discrimination can represent the result of a communal attempt to expropriate minorities. Margo (1991) documents that black schools were improved when whites felt the threat of black mobility after World War I, suggesting that there some forms of discrimination could be reasonably altered as circumstances changed. Akterlof (1974) presents a model where a caste system, such as the Jim Crow south, was an unfortunate but stable equilibrium that reflected a society wide rule where members of one clique are punished for interacting with members of a second clique.

These models certainly fit many aspects of the Jim Crow south, and they may also reflect some forms of gender-based discrimination as well. As Myrdal (1944) discussed in his classic text, integration-oriented whites were no more allowed to travel in black railcars than blacks were allowed to travel white cars. Firms proudly trumpeted their whites only policies, and the system only changed with massive legal intervention from the Federal government, which can be seen as breaking the old equilibrium with outside force. It is less clear that there was an organized conspiracy against women in the mid-20th century, that was similar to the Jim Crow system in the south, or that the legal pressure exerted by the Equal Pay Act of 1963 or the Civil Rights Act of 1964 had the same cathartic impact for women that it did for African-Americans.

But while the centralized racism view can explain many aspects of racial change, it is less clear that it can explain the changing nature of views towards African-Americans and women. In centralized discrimination models, members of the ruling clique rationally respond to incentives, and have neither negative opinions nor ill will towards either women or minorities. As such, these models have little to say about changing attitudes towards blacks and women in the workplace or at home.

Arrow's statistical discrimination model provides a model that can explain discriminatory hiring practices and beliefs. The model suggests that employers and ordinary people just have a low opinion of certain groups and these low opinions lead to discriminatory behavior. Certainly, it appears to be the case that at various times employers have held a low opinion of the competence of both blacks and women.

However, the great challenge of statistical discrimination models is that they typically also assume that people are fairly rational in their belief formation. This implies that attitudes need to be tethered to reality. Yet it is difficult to accept that there was much evidence to suggest that either women or blacks were as inept as many mid-century employers appear to have thought. In previous work (Glaeser 2005), I have focused on beliefs about malevolence, rather than competence, and pointed out that while Southern voters a century ago seem to have been convinced that African-Americans were a great threat to their safety, it was whites, not blacks, who had systematically enslaved, brutalized, sexually assaulted and even killed members of the other group. It is harder to document the error in beliefs about competence, but it seems quite likely that many people had beliefs about women and minorities that were not based on any real evidence and that bore little resemblance to the truth.

If it is true that beliefs about blacks and women systematically differed from reality, it becomes necessary to focus on theories that can generate widespread divergence between the truth and beliefs. There are at least two well-known systematic biases that can potentially generate such beliefs internally, without any external persuasion: the fundamental attribution error and self-serving biases. If the fundamental attribution error leads observers to associated negative outcomes with intrinsic personal characteristics, rather than external constraints, that individuals could readily believe that poor labor market outcomes for either blacks or women represent low levels of innate ability rather than discrimination. Self-serving biases, which lead people to prefer views that make them see themselves in a positive light, could also lead white men to have negative views of blacks and women, because such views prop up white self-esteem.

While these behavioral quirks may have contributed to negative assessment of blacks and women, there are limits to the power of these theories. For example, female belief in gender stereotypes, shown in the previous subsection, cannot be the result of self-serving biases. Here, we focus on the social formation of error.

Our critical assumption is that human beings are quite sensible to social persuasion. In the models that follow, individuals will be reasonably rational, but they will not totally discount falsely generated signals about the characteristics of out-groups. One of us followed this approach earlier but while that model focused on the political sources of biased signals, we concentrate here on biased signals in markets and at home.

On one level, the social formation of error runs against a long-standing tendency of economists to assume a high level of rationality and even accuracy in beliefs. Yet if we accept that mid-20th century white males had erroneous opinions of the ability levels of blacks and women, we must consider at least the possibility that some beliefs have little basis in reality. While our approach runs against the economist's predilection for hyper-rationality, it fully embraces the role that incentives can play in the generation of all sorts of outcomes, including incorrect beliefs.

Naturally, those incentives must battle against the incentives of listeners to learn the truth. In the political context, those incentives may be quite weak. After all, no individual voter has a strong incentive to ascertain the truth about any particular story, if the truth will only serve to make has vote a bit wiser. In the labor force context, those incentives may be quite stronger.

Moreover, we will assume that widely spread falsehoods will not persist if there is obvious evidence to the contrary. In Section IV, this fact will suggest that racial or sex-based quotas are not typically stable, while glass ceilings may be. The existing of a glass ceiling towards women (or perhaps a low, dark roof for blacks in the Jim Crow South), ensures that there is no hard evidence on how women or blacks can perform in higher positions. The absence of information allows incorrect beliefs to persist.

These models also help us to distinguish discrimination from hatred. Hatred is modeled there as a belief that an out-group is malevolent, and prone to engage in harmful behavior if they are empowered. Discrimination is a belief that an out-group is different and perhaps less capable, but not necessarily harmful or malign. Hatred leads to policies such as segregation and genocide, as in-groups attempt to shield themselves from the perceived threat. Discrimination will lead to different hiring practices and perhaps even exclusion from political decision-making. Yet policies based on discriminatory beliefs will not attempt to explicitly harm the out-group, because the out-group is not perceived as dangerous. While we might try to harm people who are perceived to be malevolent before they harm us, we have little incentive to attack people who are merely somewhat dim.

Historically, African-Americans have suffered from both discrimination and hatred. They have been perceived as being less competent, and they have also been perceived as being a threat. These beliefs were able to persist, arguably, because blacks were excluded from positions where they might do harm and kept out of jobs where they could have demonstrated ability.

Women have suffered from discrimination but not typically from hatred. The primary experience of extraordinary altruism in the lives of most men is the self-sacrificing behavior of their own mothers, which would make it hard to accept that women are somehow naturally malevolent. Indeed, many of the most profound opponents of women in the workplace, who certainly subscribe and even promulgate views about female competence, have also held up women as the fairer sex that is more generous and good-hearted than men.

It is historically rare for out-groups to be simultaneously depicted as malign and incompetent. . Indeed, such views would be counter-productive if a hate-produced is looking to generate support for policies that are harmful to the out-group. If a group is incompetent, then it is less threatening and that would mean less need to engage in defensive mechanisms. Jews, for example, have historically been depicted as both malign and powerful, which together justified the use of extreme anti-Semitic policies. The Soviet Union was depicted as an Evil Empire, which called for massive U.S. military spending. If the Soviet Union was merely an evil bumbling bureaucracy then there would have been far less need for military spending.

We now turn to three models of belief formation and bias. We begin with gender stereotypes being formed by firms selling household goods. We then turn to workplace-based belief formation and end with the formation of stereotypes within the household.

III. Corporate Investment in Gender Stereotypes

Betty Friedan's *The Feminine Mystique* depicts advertisers and magazine editors as colluding to persuade women that they will be happier in the home than in the workplace. The essential element of her worldview is that spending on consumer goods is a complement to time spent in the household. If producers want to sell more consumer goods, they also have an incentive to spend to persuade women that they will be unproductive in the workplace and happy in the home.

Throughout this paper, we will focus on beliefs about women's productivity in the workplace, and we will generally treat this productivity as purely pecuniary. This is at best a simplification, and at worst wrong. Friedan is at least as focused as beliefs about the non-pecuniary nature of work in the home and job. Such beliefs are perfectly analogous to the beliefs about female ability in the workforce that we focus on here.

In this model, we assume that women choose work hours to maximize a household utility function equal to $C_{NH} + \alpha V(T_H, C_H)$, where C_{NH} refers to non-household consumption which has a price of one, T_H refers to time spent working in the household, and C_H refers to household consumption, which is purchased at an endogenously determined price P_H. Both husband and wife have a time budget of one. The husband's earnings equal $W(H_M)A_M(1 - T_{HM})$, where H_M refers to human capital level and A_M refers to ability level and T_{HM}, refers to the amount of time that the husband spends working in the household. The function V(.,.) is assumed to be concave.

The wife's earnings will equal $\delta W(H_F)A_F(1 - T_{HF})$. The term δ is meant to capture any potential discrimination in the labor market, and H_F refers to human capital level and A_F refers to ability level and T_{HF}, refers to the amount of time that the husband spends working in the household. As long as $W(H_M)A_M > \delta W(H_F)A_F$, which we assume and that it is not optimal to have more than one person's entire life spent on household,, then it will be optimal for the wife to specialize in the household sector. We will also assume that it is not optimal for the wife to spend all of her time in the sector, so that she is on the margin between working and not-working.

If the wife's workplace-related ability is known, and we will drop that assumption shortly, the first order conditions that define an optimum are $\delta W(H_F)A_F = \alpha V_T(T_H, C_H)$, and $P_H = \alpha V_C(T_H, C_H)$, This produces our first result:

Proposition 1: Holding prices constant, time spent in the household declines with δ, H_F, and A_F and decrease with P_H if and only if $V_{CT} > 0$. Spending on household consumption will always decline with P_H and decrease with δ, H_F, and A_F and P_H if and only if $V_{CT} > 0$.

This proposition provides the core logic behinds Friedan's logic that companies selling consumer goods might benefit if women spent more time on household production, and could even potentially gain from discrimination against women in the labor market. If consumer goods and time in the household are complements, this will mean that total spending on consumer goods increases if women face stronger discrimination in the labor market, or have lower levels of human capital or higher levels of ability.

Naturally, this result is not entirely general. The quasi-linear form eliminate the possible role that income effects can play in spending on consumer goods, and means that lower levels of household income, from women not working, does not yield lower levels of spending. Moreover, in some cases, it is possible that household consumption items, like dishwashers, are a substitute rather than a complement with women's time in the workplace.

To embed this in a model of belief formation, we assume that function V(.,.) has the form $T_H^{\varphi\sigma}, C_H^{\varphi(1-\sigma)}$, where $0 < \varphi < 1$ and $0 < \sigma < 1$. With this assumption it follows that $C_H = (\alpha\varphi\sigma^{\sigma\varphi}(1-\sigma)^{1-\sigma\varphi})^{\frac{1}{1-\varphi}} P_H^{\frac{\varphi\sigma-1}{1-\varphi}} (\delta W(H_F)A_F)^{\frac{-\varphi\sigma}{1-\varphi}}$ and $T_H = \left(\alpha\varphi\sigma^{1-(1-\sigma)\varphi}(1-\sigma)^{(1-\sigma)\varphi}\right)^{\frac{1}{1-\varphi}} P_H^{\frac{-\varphi(1-\sigma)}{1-\varphi}} (\delta W(H_F)A_F)^{\frac{\varphi(1-\sigma)-1}{1-\varphi}}$. To capture the role of persuasion, we assume that there is uncertainty about A_F. If the uncertainty is not resolved before consumption and work decision are made, then this will just require replacing the actual value of A_F with its expected value. The quasi-linear utility function implies risk neutrality, and that all of the uncertainty gets taken on in the form of consumption of the numeraire good. The decision-maker sets the expected value of time in the workforce equal to the marginal utility of household time spent.

We are assuming that women must choose their hours of work before observing their ability level (we drop that assumption in Section VIII) and that women are able with probability "1-p" and in this case $A_F = 1$. They are less able with probability "p" and in this case their workforce productivity equals 1-a. Specifically, we assume that women are born believe that with probability one-half women are able in the workforce with probability $p_0 - \Delta$ and with probability one-half they have ability equal to $p_0 + \Delta$. Thus, if they had no further information they would deduce that their probability of being less able equals p_0.

In equilibrium, we assume that women observe N_G examples of women being successful in work and N_B. We also assume that women are sufficiently savvy to recognize the possibility that people are trying to persuade them, so that they believe that only N_B of the bad stories are true and the rest may be discarded as being fake. We let N_T denote the total number of believed stories. The probability that women are generally less able with probability $p_0 + \Delta$ equals $\frac{(p_0+\Delta)^{N_B}(1-p_0-\Delta)^{N_G}}{(p_0+\Delta)^{N_B}(1-p_0-\Delta)^{N_G}+(p_0-\Delta)^{N_B}(1-p_0+\Delta)^{N_G}}$ and the probability that women are generally less able with probability $p_0 - \Delta$ equals $\frac{(p_0-\Delta)^{N_B}(1-p_0+\Delta)^{N_G}}{(p_0+\Delta)^{N_B}(1-p_0-\Delta)^{N_G}+(p_0-\Delta)^{N_B}(1-p_0+\Delta)^{N_G}}$.

Stories can be both true and manufactured. We assume that the decision-maker is aware of $N_{G,0}$ examples of positive labor market outcomes for women, and that no stories of this type are manufactured. Women also understand that none of these stories are manufactured.

There are $N_{B,0}$ bad stories that are real and S_T total manufactured bad stories. We assume that women believe that $N_{B,1} + \theta(N_{B,0} + S_T)$ of the bad stories. This assumption nests a number of possible assumptions about the credulity of the listeners. If $N_{B,1} = N_{B,0}$ and $\theta = 0$, then women know exactly the number of stories that are true and added stories will have no effect on their beliefs. If $N_{B,1} = 0$ and $\theta = 1$, then listeners are credulous Bayesians, as in Glaeser and Sunstein (2008), believing everything in between.

The posterior probability that the women has low ability equals $\frac{(p_0+\Delta)^{N_{B,1}+\theta(N_{B,0}+S_T)+1}(1-p_0-\Delta)^{N_G}+(p_0-\Delta)^{N_{B,1}+\theta(N_{B,0}+S_T)+1}(1-p_0+\Delta)^{N_G}}{(p_0+\Delta)^{N_{B,1}+\theta(N_{B,0}+S_T)}(1-p_0-\Delta)^{N_G}+(p_0-\Delta)^{N_{B,1}+\theta(N_{B,0}+S_T)}(1-p_0+\Delta)^{N_G}}$ and the derivative of this with respect to S_T equals:

$$\frac{2\theta\Delta Log\left(\frac{p_0+\Delta}{p_0-\Delta}\right)(p_0+\Delta)^{N_{B,1}+\theta(N_{B,0}+S_T)}(1-p_0-\Delta)^{N_G}(p_0-\Delta)^{N_{B,1}+\theta(N_{B,0}+S_T)}(1-p_0+\Delta)^{N_G}}{\left((p_0+\Delta)^{N_{B,1}+\theta(N_{B,0}+S_T)}(1-p_0-\Delta)^{N_G}+(p_0-\Delta)^{N_{B,1}+\theta(N_{B,0}+S_T)}(1-p_0+\Delta)^{N_G}\right)^2} > 0.$$

The second derivative is negative since $\left(\frac{p_0+\Delta}{p_0-\Delta}\right)^{N_{B,1}+\theta(N_{B,0}+S_T)} > \left(\frac{1-p_0-\Delta}{1-p_0+\Delta}\right)^{N_G}$.

There are Q total suppliers of the household product in the market, who first manufacture false examples of female failure in the labor force at a cost of k. After promulgating these stories, they sell household products, engaging in Cournot competition with the other firms. All firms have a unit cost of one. If the number of women equals M, then there optimal firm behavior means that post-advertising profits will equal: $\omega M Q^{\frac{\varphi\sigma-1}{1-\varphi}}(\delta W(H_F)A_F)^{\frac{-\varphi\sigma}{1-\varphi}}$

where $\omega = \left(1-\varphi(1-\sigma)(1-\varphi\sigma)\right)(\alpha\varphi^{2-\varphi\sigma}\sigma^{\sigma\varphi}(1-\varphi\sigma)^{1-\varphi\sigma}(1-\sigma)^{2-2\sigma\varphi})^{\frac{1}{1-\varphi}}$ which is $\left(1-\varphi(1-\sigma)(1-\varphi\sigma)\right)$ times each firm's output.

Each firm has the opportunity to invest in stories, at a cost k, per story documenting some instance where a woman has entered the workforce and been unsuccessful. In practice, this may take the form of stories illustrating the bliss of staying at home. This assumption is far simpler than the relatively complicated worldview suggested by Friedan, in which magazine editors are part of a general conspiracy to promote non-working women, but it may be less accurate. The total number of stories S_T sums the investment of each individual firm.

Thus, each firm j chooses S_j to maximize $\omega M Q^{\frac{\varphi-1}{1-\varphi}}\left(\delta W(H_F)\left(1-p\left(N_{B,1}+\theta N_{B,0}+\theta \sum_{i=1}^{Q} S_i\right)a\right)\right)^{\frac{-\varphi\sigma}{1-\varphi}}-kS_j$. We assume that the second derivative of this with respect to S_i is negative near the maximum so that the first order condition characterizes a maximum.

Proposition 2: Total investment per firm is rising with M and declining with k, Q, δ and $W(H_F)$, and total market investment is falling with Q, An increase in θ causes the total number of negative stories that are believed to rise. As such, women's assessment of their ability in the workforce is rising with k, Q, δ, $W(H_F)$, and θ and rising with M.

Proposition 2 presents the comparative statics related to the belief formation by producers eager to sell consumer goods. Perhaps most notably, investment by firms is a function of market structure. Firms in highly competitive industries will not have the incentive to invest in industry-level public goods, like gender-specific norms. As such, if Friedan's model is right, then the model suggests that the market for consumer goods during her day must have been almost monopolistic in nature.

It is of course possible that some external agents managed to coordinate across disparate firms. It is possible, for instance, that the magazine editors that Friedan describes played this role. However, they would need some mechanism for solving the free rider problem, and it isn't obvious that they had that much clout.

The proposition also delivers other comparative statics that are less surprising. Higher values of k, the cost of transmission, will reduce the spread of misinformation. That effect may explain the rise of misinformation in Friedan's era when magazines had become more common and more effective. The rise of television could also have played a role in reducing the costs of persuasion. An alternative explanation is that M increased because of national markets for goods (and media). The nationalization of the market makes the returns to persuasion higher and increases the returns from persuasion.

IV. Discriminatory Belief Formation in the Workplace

We now turn to the creation of negative stereotype about women in the workplace. In this case, the stereotype is generated by members of an in-group (presumably men in the case, but it could be any group divide), who are interested primarily in keeping their jobs. I consider a system of N people who are initially peers, and at some point, a decider will be randomly selected out of the group. If the decision-maker was known in advance, then workers would focus all their persuasion on that person. Yet it seems likely that in many workplace environments, it is never entirely obvious who may play a role someone's promotion or retention. That fact generates an incentive for more widespread promulgation of discriminatory beliefs.

At the point of decision, the decider will then choose whether to replace current members of the group with outsiders. That replacement is assumed to be once-and-for-all and made solely on the basis of whether the insiders (observed) productivity is higher than the expected productivity for hiring an outsider. The decision-maker will observe the ability of the insider, but will have to base their assessment of the outsiders expected probability on the only observable characteristic of that person: their gender. These assumptions are stylized, but the basic implications of the model would generalize to a setting where there was more information about outsiders as well.

With probability π, the decider will have the opportunity to replace the internal worker with a potential replacement. The replacement may be a member of the in-group with probability ρ and that outgroup with probability $1-\rho$. The decided has observed the productivity of the insider, and receives a noisy signal of the productivity of the outsider. The decider then infers an expected productivity of the outsider and replaces the insider if and only if the outsider's productivity is higher.

Total productivity for each person combines common and idiosyncratic components: $A_i + \varepsilon_i$. Neither term is observable at the time of hire, and we consider the case in which individuals promulgate error, after learning the common component of his ability level but before learning the idiosyncratic highly job-specific component. While we refer to A_i as the more general component of ability, but nothing precludes it from being specific to this sector or firm. The ε term is distributed with mean zero, cumulative distribution F(.) and density f(.). While decision-makers observe ε, they do not observe the value of ε for prospective hires, only for incumbents.

Decision-makers are born uncertain about the distribution of productivity among the distribution of women. They initially believe that with probability one-half, the share of women that have low general ability levels, i.e. $A_i = 1 - a$, equals $p_0 - \Delta$ and the share of women with high general productivity levels, i.e. $A_i = 1$, equals $1 - p_0 + \Delta$. Decision makers also initially believe, with probability one-half, a share $p_0 + \Delta$ of women have low ability and a share $1 - p_0 - \Delta$ Thus, if they had no further information they would deduce that their probability of being less able equals p_0. Yet as they acquire examples of female achievements, they will change their assessment of the probability that a high or low share of the out-group is less able.

Decision-makers are also exposed to binary signal about general ability levels. These stories are purely a reflection of the value of A, and yield no information about ε or its distribution. Each signal is simply an observation about whether a member of the out-group has a high or low value of A_i. As opposed to the previous model, there is no falsification of stories—they are correct. However, transmitters of information do have the option to choose not to supply stories that depict the out-group in a positive light, and as such, even if the stories are true they can still be misleading.

Specifically, we assume that individuals are initially endowed with a stock of N_G examples capable out-group members and $N_{B,0}$ examples of incapable out-group members. Every person also has the opportunity to transmit q stories, at a cost c(q) to every other member of the group. Typically, we will assume that c(q) is such that q is always less than $N_{B,0}$.

The potential decision-makers are not fools, and they do recognize that their peers have a possible incentive to distort the truth by only reporting stories that make the out-group look bad. However, with probability θ, the listener forgets the source of the example and thinks that came from the original unbiased stock of stories. With probability $1-\theta$, he remembers to ignore the story because he knows that it came from an interested party. The unbiased stories are always known to be unbiased. If the decision-maker hears an additional $N_{B,S}$ stories documenting the failures of the out-group, he believes that he has N_G positive stories about the out-group and $N_{B,0}+\theta N_{B,S}$ negative stories about the out-group.

Using the same formula as the previous section, the posterior probability that the decision-maker assigns to a woman being low ability is $\frac{(p_0+\Delta)^{N_{B,0}+\theta N_{B,S}+1}(1-p_0-\Delta)^{N_G}+(p_0-\Delta)^{N_{B,0}+\theta N_{B,S}+1}(1-p_0+\Delta)^{N_G}}{(p_0+\Delta)^{N_{B,0}+\theta N_{B,S}}(1-p_0-\Delta)^{N_G}+(p_0-\Delta)^{N_{B,0}+\theta N_{B,S}}(1-p_0+\Delta)^{N_G}}$, which we denote B($\theta N_{B,S}$) . The second derivative is again negative. The posterior probability that any given potential decision-maker has that the women are less able equals: $1-aB(\theta N_{B,S})$. The decision-maker's choice, if faced with an option to replace a worker with a female alternative is to fire the male worker if and only if his total ability level is less than $1-aB(\theta N_{B,S})$. If fired, the worker loses benefits equal to W.

If the worker knows his general ability level, but not his idiosyncratic ability, at the time when he is broadcasting messages, then his probability of being replaced equals conditional upon being compared with a female alternative equals $F(1-aB(\theta N_{B,S})-A_i)$, which is the probability that $A_i+\varepsilon_i<1-aB(\theta N_{B,S})$. Since the probability that anyone he talks to will become the decision-maker is 1/Q and the probability that the decision-maker will have the option to replace him with a woman is $\rho\pi$, the overall probability that any given co-worker will replace him with a woman is $\rho\pi F(1-aB(\theta N_{B,S})-A_i)/Q$. As before, we will treat stories as a continuous variable and assume that c'(0)=0. If $N_{B,S}=N_{B,S,\neq i}+N_{B,S,i}$, where $N_{B,S,\neq i}$ represents other people's stories transmitted to this decision-maker and $N_{B,S,i}$ reflects the stories told by person i, the first order condition for the co-worker is that:

(2) $a\theta B'(\theta N_{B,S})W\rho\pi f(1-aB(\theta N_{B,S})-A_i)/Q=c'(N_{B,S,i})$.

We assume that f(.) is not too concave so that second order conditions hold. The implicit function theorem then implies that:

Proposition 3: Holding the supply of others constant, the number of negative stories told about out-group members are increasing with W, ρ and π, and decreasing with A_i if and only if

$f'\left(1 - aB\left(\theta(N_{B,S,\neq i} + N_{B,S,i})\right) - A_i\right) > 0$. The number of stories is declining with $N_{B,S,\neq i}$ as long as $f'\left(1 - aB\left(\theta(N_{B,S,\neq i} + N_{B,S,i})\right) - A_i\right)$ is not too strongly negative. The number of stories is increasing with θ if and only if $\frac{B'(\theta N_{B,S}) + \theta N_{B,S} B''(\theta N_{B,S})}{a N_{B,S} B'(\theta N_{B,S})^2} > \frac{f''(1 - aB(\theta N_{B,S}) - A_i)}{f(1 - aB(\theta N_{B,S}) - A_i)}$.

The proposition yields intuitive results. The incentive to spread malicious stories about the out-group increases with the quasi-rents associated with the job (W). As workers value keeping their jobs more highly, then are willing to use more resources to perpetuate negative stereotypes to discredit an out-group competitor. The incentive to spread stories also increases as π rises, so workers who are really sure in their job are unlikely to spread stories.

The incentive to spread stories also rises with ρ, which means that workers who are in industries where hiring a woman is not particularly likely are less likely to spread stories. As such, gender stereotypes are going to be more often spread when women are likely to be hired into the industry. In industries where men seem to have, for whatever reason, a lock on jobs, there is relatively little incentive to spread negative stories.

Higher ability workers are less likely to spread stereotypes if they are higher ability if $f'\left(1 - aB\left(\theta(N_{B,S,\neq i} + N_{B,S,i})\right) - A_i\right) < 0$. If f(.) is single peaked around zero, this suggests that workers who are innately more competent than the expected women will be are less likely to spread stories as they become more competent, but workers who are less competent than the average woman is thought to be will spread more stories as they become more competent, because this means that they are most likely to be on the margin of being replaced by a woman.

Higher values of a will typically increase the incentive to spread adverse stories about the out-group, except if changes in that value significantly reduces the probability that the individual is on the margin between being fired and not. The condition that $f'\left(1 - aB\left(\theta(N_{B,S,\neq i} + N_{B,S,i})\right) - A_i\right)$ is not too strongly negative essentially rules out that case.

Finally, the impact of θ is ambiguous. As θ changes there are three distinct effects. The first effect is that higher values of θ raise the impact of the marginal story, and that tends to make stories more prevalent. The second impact is that a higher values of θ increase the stock of adverse stories that are already believed, and that lowers the marginal impact of an extra story. These two effects are roughly akin to standard price and income effects in normal demand models. The third effect is that changes in θ may act to move the persuader away or towards from the margin of being fired, depending on the sign of $f'\left(1 - aB\left(\theta(N_{B,S,\neq i} + N_{B,S,i})\right) - A_i\right)$.

We now turn to aggregate effects assuming that individual persuaders are otherwise identical, all with values of A_i equal to one. In this case, when f(.) is symmetric and single-peaked, $f'\left(-aB\left(\theta(N_{B,S,\neq i} + N_{B,S,i})\right)\right) > 0$. To capture the impact of other workplace experience of out-group members we let $N_{T,0}$ reflect the initial supply of evidence on the out-group. We assume that $N_G = (1-p^*)N_{T,0}$ and $N_{B,0} = p^* N_{T,0}$, where $p^* \geq p_0$. Increases in female representation in the profession or firm can be thought of as increasing the value of $N_{T,0}$

The first order condition can be written as:

(2') $a\theta B'(\theta Q N_{B,S,i}) W \rho \pi f(-aB(\theta Q N_{B,S,i}))/Q = c'(N_{B,S,i})$,

and that leads to:

Proposition 4: The level of individual investment in providing examples to a single decision-maker will rise with W, ρ and π and decline with Q. Total persuasion will decline with Q if and only if $c'(N_{B,S,i}) > N_{B,S,i} c''(N_{B,S,i})$. The level of persuasion (aggregate and individual) will decline with $N_{T,0}$, as long as $p^* < \frac{Log\left(\frac{1-p_0+\Delta}{1-p_0-\Delta}\right) - \frac{\theta N_{B,S}}{N_{T,0}} Log\left(\frac{p_0+\Delta}{p_0-\Delta}\right)}{Log\left(\frac{1-p_0+\Delta}{1-p_0-\Delta}\right) + Log\left(\frac{p_0+\Delta}{p_0-\Delta}\right)}$ and $f'\left(-aB\left(\theta N_{B,S,}\right)\right)$ is sufficiently small.

As group size increases the level of individual persuasion declines. Larger group sizes mean that any one person is less likely to be the decision-maker in the future. There can be a reinforcing factor that occurs if total investment still grows because the number of persuaders grows. Higher investment by others reduces the incentive for any one individual to invest.

But total group-level investment will not necessarily decline with Q, because there are more people investing. The critical condition is that the persuasion cost function can only be moderately convex. If it were a power function (i.e. cN^z), then it would have to have a power coefficient (i.e. z) below two. For group level investment to decline we must have the individual effect, which reflects the reduced probability of any one person being the decision-maker, offset the impact of more potential persuaders. Group level investment would be more likely to decline if persuasion costs were increasing in the total number of messages sent, not just the number of messages sent per recipient.

This result speaks to the question of whether we should expect bias to be more prevalent in small or large groups. In large groups there are a large number of people spreading nonsense, and this may lead to more widespread error. However, the incentives to persuade may be stronger in small groups and this may reduce the level of bias. The question of whether bias is more or less common in urban areas seems related to the battle between these two forces.

The second result in Proposition 4 concerns the impact of $N_{T,0}$, which we think of as reflecting alternative sources of information about the out-group. If p* is sufficiently low (if $N_{T,0}$ is large, the maximum is approximately equal to $\frac{p_0-\Delta}{1-2\Delta}$, then added experience leads to a stronger belief that the out-group is competent. This added experience diminishes the marginal impact of negative stories.

There is a countervailing effect which occurs if more experience with the out-group leads potential decision-makers to have a higher assessment of the out-group and that makes the out-group more of a potential threat. That effect means that more experience with the out-group can potentially lead to more persuasion, but we eliminate that effect by assuming that $f'\left(-aB\left(\theta N_{B,S,}\right)\right)$ is sufficiently small.

This proposition speaks to the relative power of glass ceilings and quotas. If ability is very job specific, then hard glass ceilings insure that there is no world experience doing the task. This may mean that decision-makers end up being quite susceptible to stereotypes and that will ensure that steady supply of stereotyping messages. Glass ceilings therefore seem likely to be quite durable.

By contrast, quotas the limit the number of women in a profession will ensure a steady stream of women in the job providing counter-evidence to stereotypes and ultimately reducing the spread of stereotypes themselves (if the conditions in the proposition hold). This suggests that the beliefs that lead the quotas to be put in place are unlikely to persist.

This presents a similar take on the barriers discussed by Goldin (2006b), who presents evidence suggesting the discriminatory barriers at the occupation level rose and then fell over the 20^{th} century. This model can also help us to understand that fact, since men won't bother building discriminatory barriers until women become a serious threat to employment This model does not feature the same contamination effects discussed in her work, but it does suggest that men may rapidly leave an occupation once a modest number of women enter, although in this case the dynamics occur for an entirely different reason. Once there are enough women in the job, the ability and incentive to build discriminatory beliefs declines and men lose their ability to keep more able women out of the job.

V. Belief Formation in the Family: Spouses

We now turn to the creation of beliefs in the household. In some societies, perhaps America today, it may seem preposterous to think that husbands and parents would try to reduce a woman's assessment of her ability levels. We are agnostic about the question of belief manipulation in modern America, but in historic episodes where patriarchy was stronger, there seems to have been norms of parents indoctrinating daughters and especially daughters-in-law, to accept submission. In some cases, these included the view that women were unlikely to achieve success in the workforce.

We first examine the behavior of husbands and then turn to parents who inculcate beliefs in their daughters. In this case, we still focus on beliefs about women's ability in the workplace, but it may be far more relevant to consider beliefs about the relative emotional returns to working inside and outside of the home.

We now consider a two stage model of parental investment in education and beliefs, followed children's choices of marriage, time use and fertility. I will turn to a dynamic model shortly, where young adults can decide to postpone work, but initially I assume that individuals allocate a lifetime of work between childrearing and work. Both men and women are endowed with one unit of time. Each child requires a fixed time cost of $\underline{t}_C$, which I initially assume can only be borne by the wife, so her time spent working is $1 - N\underline{t}_C$, where N reflects the number of children.

Total earnings of the husband equals $W(H_M)A_M$, since the husband spends all of his time working. Total earnings of the wife equals $\delta W(H_F)A_F(1 - N\underline{t}_C)$. The term δ is meant to capture discrimination in the labor market, if it exist. The welfare of the husband equals $\theta_{MM}W(H_M)A_M + \theta_{MF}\delta W(H_F)A_F(1 - N\underline{t}_C) + \alpha V(N)$. The welfare of the wife equals $\theta_{FM}W(H_M)A_M + \theta_{FF}\delta W(H_F)A_F(1 - N\underline{t}_C) + \alpha V(N)$. The function V(N) is increasing and concave. Concavity may be interpreted as a reflection of reduction in the resources available per child as the number of children increase.

By multiplying earnings by θ_{ii} and θ_{ij} the model allows for the possibility of joint consumption of household public goods, and that spouses may get more benefit from their own earnings. For example, if $\theta_{MM} = \theta_{MF} = \theta_{FF} = \theta_{FM} = k$, where k is some constant, then all earnings are shared equally within the household. If k is equal to .5, then the two spouses just equally split all their earnings, then there are no scale economies in this form of consumption. If k=1, then all spending is a public good within the household. If $\theta_{MM} > \theta_{MF}$, then husbands are able to spend their own money on things that they prefer to buy, and likewise $\theta_{FF} > \theta_{FM}$ implies that wives too have a preference for their own earnings, perhaps because they have more control over them within household bargaining.

The term A may equal one, with probability $1 - p_i$ or $1 - a < 1$ with probability p_i, where i=M or F. These probabilities will reflect beliefs formed in the first period of the model. I first assume that childrearing choices are collective made to maximize μ times the welfare of the husband plus $1 - \mu$ times the welfare of the wife. When I allow the information of the husband and wife to differ, I assume that the welfare of each spouse is evaluated given the spouse's own knowledge at that point. Since I have assumed a potentially quite different sharing rule for income, it is quite possible that the wife's preferences dominate child-bearing, even if the man gets the lion's share of the benefits from earnings. For proposition 1, I assume that both men and women share an assessment of the probability that the women is capable, which I denote p_F.

Proposition 5: The number of children is increasing with α, p_F and a, and decreasing with $\underline{t}_C$, H_F, and δ. If $\theta_{FF} = \theta + \Delta$, and $\theta_{MF} = \theta - \Delta$, then the number of children is always decreasing with θ, decreasing with Δ if and only if $\mu < .5$., and increasing with μ if and only if $\Delta > 0$. If $\theta_{FF} > \theta_{MF}$ then the husband's welfare is increasing with the number of children and if

$(1-\mu)\left(\frac{\theta_{FF}}{\theta_{MF}}-1\right)\frac{N\underline{t}_C}{1-N\underline{t}_C} > N\frac{-V^{"}(N)}{V'(N)}$, then the husband's welfare is increasing with p_F and a and decreasing with δ and H_F.

The first part of the proposition delivers the standard results relating to benefit, time cost and opportunity cost of time. When children take more time, parents have fewer children and when the taste for children is stronger, then parents have more children. If the wife is less economically productive, then her opportunity cost of time is lower and it makes sense to have more children. An increase in male bargaining power increases the number of children if and only if the wife benefits more from her earnings than the husband benefits from her earnings. As the wife derives more benefit from her earnings, the number of children will decline if she has more bargaining power over the number of children.

Since the husband is not a dictator, he may want more children on the margin than the household actually delivers. This will occur whenever the wife derives more benefit from her earnings than the spouse. Indeed, if that condition holds, it is possible that the husbands demand for children is so strong, that he would prefer a poorer or less able wife because she will spend more time having children.

We assume that the husband can try to mislead the wife about the true extent of her workplace abilities. Assume first that if the wife is able, she receives a signal equal to one with probability π. If she is not able she receives no signal. The husband can pay a cost k, which will destroy the signal with probability g(k), which is increasing and concave, and where g(0)=0, and $lim_{k\to 0}g'(k)=\infty$.. In equilibrium, the wife believes that the husband destroys the signal with probability $\gamma g(k)$, but the husband's choice of k does not influence the wife's beliefs since it is not observed. The term γ allows for the possibility that the wife is overly trusting, or overly suspicious, of her spouse.

Thus, if the wife receives the signal, she believes with probability zero that she is less able. If the wife does not receive a signal then she believes that she is able with probability $\frac{p_F^0}{p_F^0+(1-p_F^0)(1-\pi+\pi\gamma g(k))}$, where p_F^0 refers to the prior probability of her being less able, and $\hat{\delta}$ equals her assessed probability that her husband destroyed her positive signal. I assume that optimal number of children is chosen as if the husband and wife share the same assessment of her ability, even thought the husband may know that he destroyed the signal and therefore have a higher assessment of the wife's underlying talent than she does herself. This leads to:

Proposition 6: The husband will spend positive effort to destroy the wife's positive signal of ability if $1-\frac{ap_F^0}{p_F^0+(1-p_F^0)(1-\pi)} > \frac{\theta_{MF}}{\theta_{FF}}$. If that condition holds, then the effort spent reducing the wife's assessment of her own abilities is increasing with a, π and p_F^0 and decreasing with γ. If $1-\frac{ap_F^0}{p_F^0+(1-p_F^0)(1-\pi)} > \frac{\theta_{MF}}{\theta_{FF}}$ and the third and higher derivatives of the value function, V(.) are sufficiently small so that $V''(N^+) \approx V''(N^-)$, then k is increasing with θ_{FF} and decreasing with μ, $\underline{t}_C$, H_F, and δ.

The proposition gives a condition under which the husband would like to convince his wife that she is less capable. The critical requirement is that the husband actually wants the wife to work less than the wife herself wants to work. In the model, this gap can occur if the wife's benefits more from her earnings, on a dollar by dollar, basis than the husband benefits, which could reflect a greater ability to direct money that one has made oneself, either because of physical reasons (the husband may not see the cash) or psychological ones. Naturally, there are many occasions where the husband may not want more children than the wife, and in that case, if anything the husband would like to convince the wife that she is particularly likely to be successful in the workplace.

The conflict between the spouses could occur for other reasons. I have not assumed any gap in flow utility between time spent at work and time spent in the home. If flow utility is higher at work, perhaps because home production involves particularly unpleasant labor, then the wife may have a reason to work professionally that the spouse does not internalize. Alternatively, the husband may have a particularly strong taste for home produced services that is not shared by the wife who is actually producing these services.

If this condition holds, and the husband does want to persuade the wife that she is less competent, the proposition also delivers a set of additional results. The husband's incentive to persuade increases with "a", the ability gap, and with "p", the probability that an able person doesn't produce a good signal.

VI. Influencing Sons and Daughters through Education and Maternal Lifestyle

We now move back a generation and examination persuasion by parents towards daughters and sons. To simplify concerns, we assume that $\theta_{ij} = 1$, so that men and women receive the same returns from both spouses working. They may still differ in their optimal number of children, however, if they have differences of opinion about the wife's ability in the marketplace that remain unresolved. Moreover, we assume that there is no added information revelation in adulthood, so the couple's opinions about the wife's productivity, or about women's productivity in general, are formed early on. We focus on the case where the daughter will get married, and that husbands are homogeneous.

The critical assumption is that the parents care both about the welfare of their children and about directly about their grandchildren, and specifically that the welfare of parents (that is related to a specific child) equals:

(3) $$\alpha_1\big(W(H_M)A_M + \delta W(H_F)A_F(1 - N\underline{t}_C\ ) + \alpha V(N)\big) + \alpha_2 V(N) - H_i - Other\ Costs$$

where H_i refers to the investment of human capital (for the relevant child). The function W(H) is assumed to be increasing and concave, and we assume that parents' welfare is also concave in H, including the impact of H on N, which requires that $W"(H_F)\left(\alpha_1(1-N\underline{t}_C\ )+\frac{\alpha_2 V\prime(N)\underline{t}_C}{\alpha V"(N)}\right)+\left(\frac{\alpha_2}{\alpha}-\alpha_1-\alpha_2(1-a)\frac{V\prime(N)V\prime\prime\prime(N)}{\alpha(V"(N)^2}\right)\frac{\delta A_F\left(\underline{t}_C W\prime(H_F)\right)^2}{\alpha V"(N)}$ is negative for all weakly positive values of α_2, which ensures that the first order condition $\alpha_1\delta W'(H_F)A_F(1-N\underline{t}_C\ )+\alpha_2 V'(N)\frac{\delta W\prime(H_F)A_F\underline{t}_C}{\alpha V"(N)}=1$, characterizes a maximum.

Parents care about the welfare of their son or daughter directly, and that is weighted by the constant α_1 but they also care directly about the number of grandchildren, and that is weighted by the constant α_2. This means that, given the same information, parents will always prefer their children to produce more grandchildren than the children would like to produce on their own. We do not allow conflict in preferences in the paternal generation.

We first focus on investments in a daughter's human capital, assuming that A_F is known at every point. We then turn to the possible scenario in which the parent, but not the daughter has received a private signal about the daughter's ability, in which case invest in education can serve as a costly signal to the daughter of her skills. Finally, we address sexist indoctrination of sons.

Proposition 7: Parents will invest a positive amount in daughters' education if $1>\left(1-\frac{\alpha_2 V\prime(\mathrm{N}_0)}{\alpha_1\alpha\mathrm{N}_0 V"(\mathrm{N}_0)}\right)\mathrm{N}_0\underline{t}_C$, where N_0 represents the number of children that the daughter will have if she has no education investment. If parents do invest in a positive amount of education, then the level of education is declining with α_2 and increasing with α_1. The level of education will increase with δ and A_F if and only if $\alpha_1-\frac{\alpha V"(N)}{\delta^3 A_F^2\underline{t}_C W\prime(H_F)W(H_F)}>\frac{\alpha_2}{\alpha}\left(1+\frac{V\prime\prime\prime(N)V\prime(N)}{(V"(N))^2}\right)$

Now we allow parents to know their daughters ability, but the daughters themselves can only infer their talents from their parental investment in their human capital. A fraction of parents denotedp_F now that their daughter has productivity 1 and the remainder know that their daughter has productivity 1-a. The parents cannot convincingly transmit their knowledge to their daughters, but their investment in human capital will serve as a signal.

I define $N_S(H)$ by $\delta W(H)\underline{t}_C=\alpha V'(\ N_S(H))$, and $N_U(H)$ by $(1-a)\delta W(H)\underline{t}_C=\alpha V'(\ N_U(H))$, These are the fertility levels implied by human capital level H and the belief of the daughter that she is skilled and unskilled respectively. It is always true that $N_S(H)<N_U(H)$. Parents always prefer more children in equilibrium because the children who make the decision don't internalize that impact that their fertility has on parents' welfare through α_2. I go further and assume that parents of skilled children typically want to convince their daughters that they are less able, at least in the workplace environment. This will lead the parents of skilled daughters to want to imitate the parents of less skilled daughters, and it creates a signaling game involving parental investment in human capital.

There cannot be multiple levels of human capital chosen by both skilled and unskilled parents. The levels of fertility that would make the unskilled parents indifferent between the two education levels would always lead the more skilled parents to strictly prefer the more educated point. Moreover, generically, it is impossible to have multiple investment levels involving only one group of parents investing, as they will not yield the same level of welfare. As such, there are at most three levels of human capital that are being chosen—one perhaps chosen by parents of unskilled parents alone, one chosen by skilled parents alone, and one chosen by a mixture between the two types of parents.

The investment level chosen by skilled parents alone will be denoted H_F^{Skill} with fertility level N_{Skill}, if such an outcome exists, where those quantities are defined by $\delta W\left(H_F^{Skill}\right)\underline{t}_C = \alpha V'(N_{Skill})$, and $1 = W'\left(H_F^{Skill}\right)\delta A_F \alpha_1 \left(1 - \left(1 - \frac{\alpha_2 V'(N_{Skill})}{\alpha_1 \alpha N V''(N_{Skill})}\right) N_{Skill}\underline{t}_C\right)$, and generate welfare for the skilled parents denoted U_{Skill}^{Skill}. Essentially, if parents are known to have skilled daughters, then they might as well choose the optimal level of investment given that knowledge by their children. I comparably define $H_F^{Unskill} < H_F^{Skill}$ as the optimizing level of human capital by parents of unskilled parents who are known to have unskilled children. I also assume a minimum level of investment that parents are legally required to make denoted $\underline{H_F}$, which is less than $H_F^{Unskill}$.

I define $\widehat{H}$, as the lowest level of investment that makes the parents of skilled children indifferent between delivering their own ideal level of education, and being known as having skilled parents, and I assume that at H=0, the parents of skilled daughters strictly prefer H_F^{Skill}. The term $\widehat{H}$ satisfies

$$(4)\ \alpha_1\left(\delta W\left(H_F^{Skill}\right)\left(1 - N_S\left(H_F^{Skill}\right)\underline{t}_C\right) + \alpha V\left(N_S\left(H_F^{Skill}\right)\right)\right) + \alpha_2 V\left(N_S\left(H_F^{Skill}\right)\right) - H_F^{Skill} = \alpha_1\left(\delta W\left(\widehat{H}\right)\left(1 - N_U\left(\widehat{H}\right)\underline{t}_C\right) + \alpha V\left(N_U\left(\widehat{H}\right)\right)\right) + \alpha_2 V\left(N_U\left(\widehat{H}\right)\right) - \widehat{H}$$

This leads to:

Lemma 1: There exists at least one value of H, at which parents of skilled daughters are indifferent between choosing H and appearing to be the parents of unskilled daughters and choosing H_F^{Skill}. At the least value of H that satisfies this indifference condition, the welfare of parents of skilled daughters is strictly increasing in H, and this value of H is rising with δ and α_1 and falling with α_2.

To generate a single equilibrium for any given set of parameter values, we now assume a variant of the D1 refinement (Cho and Kreps, 1987) which requires that if an off-the-equilibrium path investment level is more attractive for one type of parent, given any set of beliefs by children, then children assume that this type of parent has generated this deviation with probability one. In this model, the children's response to the parents' human capital investment is their fertility level. So this assumption means that if $N_S^*(H)$ and $N_U^*(H)$ makes the parents of skilled and unskilled children respectively indifferent between their equilibrium payoff and any deviation H, then if $N_S^*(H) > N_U^*(H)$ the deviation is more likely to have come from a parent of a skilled child. If $N_S^*(H) < N_U^*(H)$ then the deviation is more likely to have come from a parent of an unskilled child. Thus the child will think that the deviation comes, with probability one, from the parent of an unskilled child if and only if $N_S^*(H) < N_U^*(H)$.

This assumption leads to proposition 8:

Proposition 8: If $\widehat{H}>H_F^{Unskill}$, then there is no pooling equilibrium, skilled parents choose to invest H_F^{Skill} and unskilled parents choose $H_F^{Unskill}$. If $H_F^{Unskill} > \widehat{H} > \underline{H_F}$, then skilled parents choose H_F^{Skill} and unskilled parents choose $\widehat{H}$. If $\widehat{H} < \underline{H_F}$, then both all unskilled and some skilled parents choose $\underline{H_F}$. At that equilibrium, the number of parents of skilled daughters choosing $\underline{H_F}$ will decrease with α_1 and δ and increase with α_2.

The proposition shows that given the equilibrium refinement, pooling only occurs when the minimum level of education binds. In that case, there is either a pure pooling equilibrium where all parents choose the minimum or a semi-pooling equilibrium where the parents of skilled daughters either choose the optimal level of schooling for them or imitate the parents of the less skilled and give their daughters the minimum schooling level.

The degree of mixing depends on the parameters. There will be more parents educating skilled daughters if discrimination in the labor market is weaker or if they are more altruistic to their daughters relative to their grandchildren.

For higher value of $\widehat{H}$, which, as described above, is determined by the altruism of the parents and the level of labor market discrimination, there is a pure separating equilibrium. The parents of the skilled go ahead and educate their daughters to a relatively high degree (although even they reduce their schooling investment in order to encourage more child-bearing). The parents of the less skilled choose a particularly low level of schooling. That low level is found so that they are not imitated by the parents of the more skilled. Eventually, if there is sufficiently little labor market discrimination or if grandparents are sufficiently disinterested in the number of grandchildren, then parents of skilled and unskilled alike, choose to invest in the level they would choose in the absence of signaling concerns.

The model has several implications. When labor market discrimination is strong, then parents of skilled and unskilled daughters alike choose to provide them with minimal education. The skilled daughters may particularly suffer, because their parents are trying to ensure that they don't realize their skills. As women are less discriminated against, this leads to more investment in the skilled daughters, and there can be a discrete jump for them. Eventually, signaling concerns vanish in a pure separating equilibrium, and the skills essentially serve to maximize welfare as described in Proposition 8.

This model suggested that the population would have only two levels of education for women, but that would not be the case if there were visible differences in parameters across the population. In that case, different parameters will lead to different equilibria, although for any given set of observable parameters, parents will still use education to influence their daughters' beliefs.

Maternal Choice of Lifestyle

We now turn from influencing daughters to influencing sons through the maternal choice of lifestyle. This lifestyle choice will also influence the beliefs of daughters, but we will focus on the persuasion of the son. The results for daughters are equivalent.

Belief manipulation requires taking actions while the boy is young in order to persuade him that his future wife's time is better spent in the home than at work. Following the assumptions of the previous sections, and assuming that the beliefs of the sons' prospective wives about their likelihood of being less able are fixed at p_W, the first order condition for fertility for the son, condition on marriage, will equal:

$$(5)\ (1-\mu p_S a-(1-\mu)p_F a)\delta W_F \underline{t}_C = \alpha V'(N).$$

where p_S reflects the sons belief that his wife will be less effective in the workforce. This term then defines $N(p_S)$, the number of children that a son has condition upon his belief about the women's ability. Differentiation reveals that the number of children is decreasing in the son's assessment of women's competence.

While this belief will, in reality, surely be influenced by particular features of the wife, we will assume here that his opinion is based only on his general belief about the workplace competence of women. This will create the most scope for parental influence on sexist beliefs, and the strongest incentive to invest in those beliefs. In the next section, we will drop this assumption and allow there to be added sources of information about the competence of the wife.

Specifically, we assume that the son's belief about the appropriate role of women is going to be shaped by the actions of his mother during his childhood. Specifically, we assume that sons are born with uniform priors about the competence of women, which are distributed between $p_0-\Delta$ and $p_0+\Delta$. Thus, if he has no further information, he will assume that the probability his wife is less competent is p_0.

He will however observe his mother's work behavior and make an inference about whether she is more or less competent. That inference will then shape his assumption about the prospective competence of all women.

Bayes' rule implies that if the son believes that his mother is less capable, his posterior probability of a random women being less capable equals $p_0+\frac{\Delta^2}{3p_0}$. If he knows that his mother is capable, his posterior probability that women are less capable equals $p_0 - \frac{\Delta^2}{3(1-p_0)}$. If he has not learned whether his mother is capable or not, he retains his prior probability of women being competent that is equal to p_0. If the child believes that the parents is less skilled with probability "p" then the posterior belief about probabilities in the population, equals $p_0 + \frac{\Delta^2(p-p_0)}{3(1-p_0)p_0}$. We assume that p_0 is the correct population probability of skilled mothers.

We also assume that mothers of this older generation have θ units of time to allocate (post childbearing) between working in the marketplace and working in the household. Household benefits are denoted H, and these are weighed against the workplace returns for the mother which equal either δW_F or $\delta W_F(1-a)$. Mothers are assumed to know their ability by this point in time and we assume that $\delta W_F > H > \delta W_F(1-a)$, so in the absence other considerations, mothers with more workplace related abilities will work while mothers with less workplace related ability will remain in the home.

If the son believed that his mothers' workplace choices were non-strategic, he would then infer from the choice or work or not, the mother's workplace ability and update his beliefs accordingly.

But of course, the parents' decisions are strategic and meant to influence the child's beliefs. Let

(6) $\Delta(p,\alpha_2) = (\alpha_1 \alpha + \alpha_2)\left(V\left(N\left(p_0 + \frac{\Delta^2(p-p_0)}{3(1-p_0)p_0}\right)\right) - V\left(N\left(p_0 - \frac{\Delta^2}{3(1-p_0)}\right)\right)\right) - \alpha_1 \delta W_F(1 - p_P a)\underline{t}_C\left(N\left(p_0 + \frac{\Delta^2(p-p_0)}{3(1-p_0)p_0}\right) - N\left(p_0 - \frac{\Delta^2}{3(1-p_0)}\right)\right)$

denote the benefits in terms of the next generations' behavior of pooling with the less skilled parents, as a function of "p", which reflects the proportion of unskilled parents in the pooling equilibrium, which can therefore range from p_0 to 1. The term p_P refers to the probability that an able parent places on his female child or his male child's wife being able in the workplace (which may perhaps equal $1 - p_0 + \frac{\Delta^2}{3(1-p_0)}$ if parents share the beliefs that children would have if they knew their parents ability). We suppress the α_2 argument until we discuss children's incorrect beliefs about parental optimism.

The derivative of $\Delta(p,\alpha_2)$ with respect to p is always positive if and only if $\frac{\alpha_2}{\alpha\alpha_1} > \frac{\mu\left(p_0+\frac{\Delta^2(p-p_0)}{3(1-p_0)p_0}\right)a+(1-\mu)p_F a - p_P a}{1-\mu\left(p_0+\frac{\Delta^2(p-p_0)}{3(1-p_0)p_0}\right)a-(1-\mu)p_F a}$, which if $p_P = p_0 - \frac{\Delta^2}{3(1-p_0)}$ and $p_P = p_0$, converges to $\frac{\alpha_2}{\alpha\alpha_1} > \frac{\Delta^2(\mu p+(1-\mu)p_0)a}{3(1-p_0)p_0(1-ap_0)-\mu a\Delta^2(p-p_0)}$, which we assume holds and we also assume that $\Delta(p_0,\alpha_2) > 0$.

The following proposition then follows:

Proposition 9: If we assume the D1 refinement, then there is a unique equilibrium at all parameter values, and if $\theta(\delta W_F - H) > \Delta(1)$, then a pure separating equilibrium exists where more workplace-able mothers' work and less workplace-able mothers don't work, if $\theta(\delta W_F - H) < \Delta(p_0)$, then there exists a pure pooling equilibrium where all mothers don't work, and if $\Delta(1) > \theta(\delta W_F - H) > \Delta(p_0)$,, then there is a semi-pooling equilibrium where a fraction of workplace-able mothers work in the market and the remained work at home. The average posterior belief of children equals p_0 in the pooling and separating equilibria, and also in the semi-pooling equilibrium, as long as children believe that the probability that their parent is less able is equal to the share of not-working parents who are less able. .

Proposition 9 characterizes a standard signaling game, where there are three parameter value ranges. If more able mothers gain greater by working, then they will work, and their children will think that they are more able while the children of less able parents will also correctly infer their parents' ability level. The children of working mothers will have too high an assessment of the average woman's ability while the children of non-working mothers will have too low an assessment of the average woman's ability.

When the incentives to work are low, for the more able parents, then everyone will not work. In that case, children don't update their assessments and keep their priors, which we have assumed are actually correct. Finally, for an intermediate range of values some of the more able parents will choose to imitate the less able by not working, this will lead to a distortion of beliefs for the two groups, but on average they two groups will get it right.

This model so far does predict that more able women will undertake some costly signaling. They will want to persuade both sons and daughters that women are better off outside of the workplace, and that will lead to too little working. However, this will not explain a society wide error on the average competence of women.

In order to potentially explain society-wide error, we will need to introduce some irrationality into belief formation. Specifically, we will follow Glaeser and Sunstein (2008) and assume that the children are credulous Bayesians, meaning that they will have too little cynicism about their parents' motives. In this case, we operationalize that by assuming that children underestimate their parents' propensity to engage in strategic behavior. One natural means of achieving that end is that assume that children underestimate the extent to which parents care about anything other than their own welfare, specifically by assuming that they assume that the parents altruism towards grandchildren equals $\hat{\alpha}_2 < \alpha_2$, which then determines a value $\Delta(p, \hat{\alpha}_2)$, which is strictly increasing in $\hat{\alpha}_2$.

Lemma 2: The function $\Delta(p, \hat{\alpha}_2)$ is less than zero when $\hat{\alpha}_2 = 0$, greater than zero (when p lies between p_0 and 1) when $\hat{\alpha}_2 = \alpha_2$, and strictly increasing in $\hat{\alpha}_2$, and whenever $\Delta(p, \hat{\alpha}_2)$ is positive, the function is strictly increasing in p between p_0 and 1.

We now define our equilibrium modified equilibrium constant:

Modified Bayesian Nash Equilibrium: A modified Bayesian Nash equilibrium will exist if and only if parents are choosing work optimally given the beliefs of their children, and children form their beliefs of parental ability by assuming that parents are behaving optimally given children's beliefs and their incorrect assessment of α_2, as such children's assessment of parental behavior may be incorrect.

This definition leads to:

Proposition 10: If $\theta(\delta W_F - H) > \Delta(1, \alpha_2)$, then the separating equilibrium exists and if we assume the D1 refinement then it is unique and children have the correct beliefs about parents' ability If $\Delta(1, \alpha_2) > \theta(\delta W_F - H)$, then for all values of $\hat{\alpha}_2$ below α_2, all parents choose not to work, but children believe that some or all skilled parents work.

If but children believe that the beliefs that children think would make skilled parents indifferent between working and not working, actually make parents strictly prefer not working, and as such all parents choose not to work.

Define $\hat{\alpha}_2^*$ as the value of $\hat{\alpha}_2$, which satisfies $\Delta(1, \hat{\alpha}_2^*) = \theta(\delta W_F - H)$ and define $\hat{\alpha}_2^{**}$ as the value of $\hat{\alpha}_2$, which satisfies $\Delta(p_0, \hat{\alpha}_2^{**}) = \theta(\delta W_F - H)$. Lemma 2 implies that $\hat{\alpha}_2^{**} > \hat{\alpha}_2^*$.

Proposition 11: When $\Delta(1, \alpha_2) > \theta(\delta W_F - H)$, and $\hat{\alpha}_2$ is less than $\hat{\alpha}_2^*$, then children believe that all skilled parents work which means that the universal posterior is that women are less able with probability $p_0 + \frac{\Delta^2}{3p_0}$. If $\theta(\delta W_F - H) < \Delta(p_0, \alpha_2)$, and $\hat{\alpha}_2$ is greater than $\hat{\alpha}_2^{**}$, children correctly know that they are in a pure pooling equilibrium. When $\Delta(1, \alpha_2) > \theta(\delta W_F - H)$, $\hat{\alpha}_2$ is greater than $\hat{\alpha}_2^*$, and either $\theta(\delta W_F - H) > \Delta(p_0, \alpha_2)$ or $\hat{\alpha}_2$ is less than $\hat{\alpha}_2^{**}$, then children incorrectly believe that some skilled parents work. The values of $\hat{\alpha}_2^*$ and $\hat{\alpha}_2^{**}$, and the perceived likelihood that a parent who is not working is unskilled in any perceived semi-pooling equilibrium must be increasing with θ and decreasing with H.

Proposition 11 delivers the core results of the model. When children overestimate parental altruism, then when $\Delta(1, \alpha_2) > \theta(\delta W_F - H)$, children believe that skilled mothers will work and only unskilled mothers stay home. But in reality, skilled and unskilled mothers alike stay home. The result is that universally children believe that their mothers are unskilled and this leads them to increase their posterior assessment that their future wives will be unskilled as well.

We have not considered the possibility that children also look at the friends' parents and infer the distribution of ability from their choices as well. This learning process would only exacerbate the results. If all mothers everywhere choose not to work, then children might infer that all mothers are less then competent. This would mean that their posterior would be that all women are less able.

An Aside on Homophobia

The core assumption of this aspect of the model is that parents want more grandchildren than their sons and daughters naturally will give to them. The same parental preferences should also generate incentives to engage in other forms of belief investment, most notably inculcating opposition to homosexual lifestyles. If homosexuality leads to less own grandchildren, then parents who value own grandchildren will invest in their children's beliefs to that end. They will attempt to convince them that homosexuality will lead to unhappiness and perhaps worse.

This fact may remind us that religious organization may offer parents a means of perpetuating beliefs that serve their biological interests. If the church supports traditional lifestyles and opposes homosexuality, then parents may have an incentive to take their children to church despite their own private religious beliefs.

VII. Timing of Work and Childbearing

In this last section, we examine the impact that changing technologies of child-rearing, and examine the impact this has on beliefs about female ability We assume that people live for a total of one period of time, and families maximize $\int_{t=0}^{1} y(t)dt + \alpha V(N)$, where y(t) represents earnings in time t, and N represents the total number of children over the lifecycle. We essentially normalize the interest rate to zero, to eliminate the comparison of timing children versus timing earnings. Since the value function for children admits no role for having children earlier in the lifespan, for balance it makes sense to simultaneously assume that there is little gain to having earnings earlier in the lifecycle. We also assume that $\alpha V(N)= \alpha N^{\mu}$.

We normalize the husband's lifetime earnings to zero, and essentially exclude him from the rest of the discussion. The wife has a production function for children that assume that if a woman begins the child-bearing process at time t_0, and spends T units of time, then the number of children produced will be $g(t_0)T/\underline{t}_C$ where the number of children produced is decreasing in the age of first child birth and increasing the total time spent in childbirth. The assumption that N is decreasing in the age of initiating childbearing attempts to capture the added costs of having children at older ages, although, importantly, that will be a function of changes in technology that we will explicitly model.

We make a restrictive assumption that childbearing occurs in a single time cluster and precludes all other work for that time period. We think of the assumption of a single childbearing period as reflecting possibly deeper assumptions about the natural of home production with children—the scale economies that can exist when multiple children are at home for example—and the fact that when human capital is built up over time, there is little advantage to interspersing periods of work amidst period of being out of work. The assumption that child-rearing is full-time is clearly an approximation, but it makes the structure of the model somewhat easier. In some cases, we may be treating a woman who is working half-time until her child is ten as a woman who is out of the labor force entirely until her child is five and working full time after that period.

We slightly permute the assumptions about wages found earlier. We assume that all female workers initially earn $(1-a)\delta W_0$, which is essentially the low skilled wage discussed previously. After a period of ψ continuous units in the labor force, the women will learn whether she is high productivity or low productivity. If she is low productivity, she will continue to earn $(1-a)\delta W_0$ in perpetuity. If she is high productivity, she will then earn δW_0 as long as she remains in the labor force continuously. If she leaves the labor force, she will again earn $(1-a)\delta W_0$ upon her return, but she will then know that she is productive and that when she reenters, she will earn δW_0 after ψ units of time.

We assume that the initial probability that she places on being less able equals p_0. We will subsequently assume that this in endogenously determined through persuasion of various forms.

As child-bearing is a continuous block of time, it can come at the beginning, middle or end of working life, but there are really only two key cases to compare. The first case represents child-bearing in ignorance, meaning that the woman enters into child-bearing before she has spent ψ units of time in the workforce, and as such does not know her ability level. The second case is child-bearing given knowledge, which represents the case that occurs if the woman starts her child-bearing after learning her ability level.

The next proposition details when the conditions that lead to the optimality of one or the other of those two options.

Proposition 12: If $\left(\frac{t_C}{\alpha}\right)^{\frac{\mu}{1-\mu}}(\delta W_0)^{\frac{1}{1-\mu}}$ is sufficiently small (when $\mu=.5$, $\left(\frac{t_C}{\alpha}\right)(\delta W_0)^2<\frac{1}{4\psi}\left(1-\frac{ap_0}{(1-a)(1-ap_0)}\right)$), then there exists a value $g(\psi)$ denoted $g(\psi)^*$ such that women are indifferent between having children immediately and working until $t_0=\psi$. For all values of $g(\psi)$ above $g(\psi)^*$, the optimal strategy is to wait until $t_0=\psi$ and for all values of $g(\psi)$ below $g(\psi)^*$, the optimal strategy is to have children immediately. The value of $g(\psi)^*$ is rising with $\underline{t}_C$, δ and $W_0$and falling with α. The value of $g(\psi)^*$ rises with a if p_0 is close to zero or one, and falls with p_0, if $\mu\leq .5$ or if p_0 is near to one.

This proposition illustrates the two basic strategies that this framework allots for mothers who decide to have children. Decision-makers either commence childbearing immediately, and then potentially switch to work after having a child, or they enter the work force, laboring until the point in which they have determined their core ability level. After that point, they immediately switch towards child-bearing. In a sense, this represents waiting until a tenure decision until having children, although relatively view tenured professors then leave the workforce altogether, and some take little time off at all, an option that the model doesn't consider.

There are two forces pushing towards earlier child-bearing: (1) there are costs to interrupting a career, since the career clock essentially begins ticking after the person learns their ability level and (2) there are biological costs that increase the cost of delay. There is one key advantage to waiting—by acquiring information about ability, the individual can make better decisions about how much time to allocate to child-bearing or not.

It is possible that even without the biological costs, individuals prefer childbearing immediately. The key condition that eliminates that possibility is that $\left(\frac{t_C}{\alpha}\right)^{\frac{\mu}{1-\mu}} (\delta W_0)^{\frac{1}{1-\mu}}$ is sufficiently small, and that condition becomes $4\psi \left(\frac{t_C}{\alpha}\right) (\delta W_0)^2 < 1 - \frac{ap_0}{(1-a)(1-ap_0)}$, when $\mu = .5$. This condition ensure that the costs of disrupting careers are not so high, so that even if there were no biological costs to waiting, people would still want to commence their careers immediately. When this condition holds, then waiting becomes a plausible option, and there is a cutoff point, based on the biological costs of waiting, at which it makes sense to move from the immediate kids strategy to the wait-and-see strategy.

The value of $g(\psi)^*$ denotes the highest possible biological cost of waiting at which it still is optimal to delay child-bearing. Higher values of $g(\psi)^*$ indicate that the threshold for waiting is higher; lower values indicate that the threshold is lower and there should be more young women in the labor force. Obviously, one implication of this is that changes in medical technology that make it easier to delay childbearing should increase the number of young women in the labor force.

The proposition also suggests that this threshold will be lower, when α is higher, and when $\underline{t}_C$, δ and $W_0$are lower. Waiting becomes more attractive when α is higher, because higher α translate into more children, and also a greater response in the number of children to better information. If the prospective parent isn't planning on having more than one child, regardless of her type, then it is less relevant to learn whether one should expect a good or bad job market career. The impact of $\underline{t}_C$ is similar. When the costs of childbearing are higher, then the threshold is higher, since information is less valuable.

Increases in δ and W_0 have two complementary effects. First, they decrease the number of desired children and the response of those children to new information. That reduces the value of waiting. Moreover, the financial cost of waiting increases as δ and W_0 rise, because of the lost earnings.

Changes in "a" and "p_0" have myriad effects. As long as the value function for children is sufficiently concave, then higher values of p_0 lower the threshold, because they increase the gap between the two states.

We know allow the possibility for persuasion about p_0. We assume that the persuading has some benefit associated with T, denoted W(T) which is a concave function, either coming from consumer spending or direct desire for more children, and that $p_0 = p_0$(e), where expenditures increase p_0. The persuader's welfare function is $W(T(p_0(e))) - e$.

This leads to proposition 13:

Proposition 13: If the decision-maker faces a value of $g(\psi)$ greater than $g(\psi)^*$, then marginal investments in persuasion will create no benefit for the persuader. If the decision-maker faces a value of $g(\psi)$ below $g(\psi)^*$, then there will be benefits from persuasion, and if the decision-maker is on the margin between waiting or not, then marginal persuasion will increase the number of children by pushing the women towards having children immediately, only if increase in $p_0$raise $g(\psi)^*$, and if the women significantly underestimates her probability of being able to being with.

This proposition helps to make sense of the time series that appears to occur in belief formation. As Goldin (2006a) describes, women were initially prone to work after marriage and then the pattern switched and more women worked earlier. That switch should, if the model's assumptions are correct, act to reduce the incentive to invest in gender-related beliefs and stereotypes. If women are waiting to learn their type before having children, then they are likely to be less responsive to parental misinformation about their ability level or likelihood of enjoying work.

VIII. Conclusion

This paper has produced a series of model which attempt to capture the different possible sources of gender stereotypes. We have modeled belief formation by consumer goods companies, co-workers, husbands and parents. We are agnostic about the relative importance of these forces at this time, but we hope that the models have generated enough testable implication to help sort out the relative importance of the different possible belief creators.

There various sections did however make a central point. The power of persuasion to indoctrinate error is limited by real world experience. If decision-makers have plenty of experience of competent women, then attempts to portray women as generally incompetent will have little effect. If women themselves have experienced work and seen their own ability level, then persuasion about gender roles will also be less effective. The changes in technology that brought women into the workforce and enabled them to postpone child-bearing may not only have enabled women to enjoy more productive work lives, but these may also have severely reduced the spread of error about women.

References

Arrow, Kenneth J. "The Theory of Discrimination." In Orley Ashenfelter and Albert Rees, eds. *Discrimination in Labor Markets.* Princeton, N.J.: Princeton University Press, 1973. 3-33.

Akerlof , George. “The Economics of Caste and of the Rat Race and Other Woeful Tales.” The Quarterly Journal of Economics, Vol. 90, No. 4, 599-617. Nov., 1976.

Becker, Gary. *The Economics of Discrimination.* Chicago: University of Chicago Press, 1957.

Blau, Francine, Peter Brummund and Albert Yung-Hsu Liu. “Trends in Occupational Segregation by Gender 1970-2009: Adjusting for the Impact of Changes in the Occupational Coding System,” *Demography* (forthcoming).

Friedan, Betty. *The Feminine Mystique.* New York: W.W. Norton. 1963.

Glaeser , Edward L. “The Political Economy of Hatred,” *Quarterly Journal of Economics*, 120(1) (2005): 45-86.

Glaeser, Edward L. and Cass Sunstein. “Extremism and Social Learning,” *Journal of Legal Analysis*, (2008).

Goldin, Claudia. *Understanding the Gender Gap.* (Oxford: OUP, 1990).

Goldin Claudia. 2006a. “The quiet revolution that transformed women’s employment, education, and family.” *American Economic Review* 96:1–21

Goldin Claudia. “The Rising (and then Declining) Significance of Gender.” In: Blau F, Brinton M, Grusky D, eds, *The Declining Significance of Gender?* Russell Sage Foundation Press; 2006b. p. 67-101.

Gunderson, Elizabeth, Gerardo Ramirez, Susan Levine, and Sian Beilock. 2012. "The Role of Parents and Teachers in the Development of Gender-Related Math Attitudes." *Sex Roles* 66, no. 3/4: 153.

Krueger, Anne O. “The Economics of Discrimination.” *Journal of Political Economy* , Vol. 71, No. 5 (Oct., 1963), pp. 481-486

Gerda Lerner. *The Creation of Patriarchy*. (Oxford: OUP, 1987).

Margo, Robert A, 1991. "Segregated Schools and the Mobility Hypothesis: A Model of Local Government Discrimination," *The Quarterly Journal of Economics*, MIT Press, vol. 106(1), pages 61-73, February.

Myrdal, Gunnar. *An American Dilemma: The Negro Problem and Modern Democracy*. (New York: Harper & Brothers, 1944).

Simon, Rita J., and Jean M. Landis. 1989. "A Report: Women's and Men's Attitudes about a Woman's Place and Role." *The Public Opinion Quarterly* 53(2) (Summer) pp. 265-276.

Thurow, Lester C. *Poverty and Discrimination* (Brookings Institution:1969).

Proofs of Propositions:

Proof of Proposition 1: Consider the two relevant first order conditions $\delta W(H_F)A_F = \alpha V_T(T_H, C_H)$, and $P_H = \alpha V_C(T_H, C_H)$. Let Z denote $\delta W(H_F)A_F$, and then differentiate both conditions totally with respect to Z. This produces the derivative $\frac{\partial T_H}{\partial Z} = \frac{V_{CC}}{\alpha(V_{CC}V_{TT} - V_{CT}^{@})}$ and $\frac{\partial C_H}{\partial Z} = \frac{-V_{CT}}{\alpha(V_{CC}V_{TT} - V_{CT}^{@})}$. The first term is negative, proving the results for δ, H_F, and A_F which increase Z and have no other impact on the conditions. The second term is negative if and only if $V_{CT} > 0$. Differentiating with respect to P_H yields: $\frac{\partial T_H}{\partial P_H} = \frac{-V_{CT}}{\alpha(V_{CC}V_{TT} - V_{CT}^{@})}$ and $\frac{\partial C_H}{\partial P_H} = \frac{V_{YT}}{\alpha^2(V_{CC}V_{TT} - V_{CT}^{@})}$. The second term is unambiguously negative and the first term is negative if and only if $V_{CT} > 0$.

Proof of Proposition 2: The firm's first order condition for investing in misleading stories about women in the workforce is:

$$\omega MQ^{\frac{\varphi\sigma-1}{1-\varphi}} \frac{\varphi\sigma}{1-\varphi} \left(\delta W(H_F)\right)^{\frac{-\varphi\sigma}{1-\varphi}} \left(1 - p\left(N_{B,1} + \theta N_{B,0} + \theta \sum_{i=1}^{Q} S_i\right) a\right)^{\frac{\varphi(1-\sigma)-1}{1-\varphi}} a\theta p'\left(N_{B,1} + \theta N_{B,0} + \theta \sum_{i=1}^{Q} S_i\right) = k$$

The second derivative is
$\omega MQ^{\frac{\varphi-1}{1-\varphi}} \frac{\varphi\sigma}{1-\varphi} \left(\delta W(H_F)\right)^{\frac{-\varphi\sigma}{1-\varphi}} a\left(1 - p\left(N_{B,1} + \theta N_{B,0} + \theta \sum_{i=1}^{Q} S_i\right)a\right)^{\frac{\varphi(2-\sigma)-2}{1-\varphi}}$ times

$\frac{1-\varphi(1-\sigma)}{1-\varphi}\left(\theta p'\left(N_{B,1} + \theta N_{B,0} + \theta \sum_{i=1}^{Q} S_i\right)\right)^2 + \left(1 - p\left(N_{B,1} + \theta N_{B,0} + \theta \sum_{i=1}^{Q} S_i\right)\right)\theta p"\left(N_{B,1} + \theta N_{B,0} + \theta \sum_{i=1}^{Q} S_i\right)$ and we have assumed that this is negative, which requires that the second term is larger in absolute value than the first term.

This expression be rewritten as: $h_S(\theta QS_i, Z) = k$ where Z represents a vector of exogenous characteristics. If $h_{SS}(\theta QS_i, Z) < 0$, which we assume, then S_i is declining with Z. For any other exogenous characteristic, the derivative of S_i with respect to Z is positive if and only if $h_{SZ}(\theta QS_i, Z) > 0$. This means that stories are declining with Q, δ and $W(H_F)$ and rising with M. The total number of stories QS_i must also be declining with Q, because $\theta h_{SS}(\theta QS_i, Z)(Q\frac{\partial S_i}{\partial Q} + S_i) + h_{SQ}(\theta QS_i, Z) = 0$, or $\frac{\partial S_i}{\partial Q}Q + S_i = -\frac{h_{SQ}(\theta QS_i, Z)}{\theta h_{SS}(\theta QS_i, Z)} < 0$.

The derivative with respect to θ satisfies $Qh_{SS}(\theta QS_i, Z)(\theta\frac{\partial S_i}{\partial \theta} + S_i) + h_{S\theta}(\theta QS_i, Z) = 0$, where $h_{S\theta} > 0$. As such $\theta\frac{\partial S_i}{\partial \theta} + S_i > 0$, or the total number of messages heard must be rising with θ.

Proof of Proposition 3: Differentiating $a\theta B'(\theta N_{B,S})W\rho\pi f(1-aB(\theta(N_{B,S,\neq i}+N_{B,S,i}))-A_i)/Q-c'(N_{B,S,i})=0$, totally with respect to any exogenous parameter Z yields $\frac{\partial N_{B,S,i}}{\partial Z}\frac{\partial}{\partial N_{B,S,i}}\left(a\theta B'(\theta(N_{B,S,\neq i}+N_{B,S,i}))W\rho\pi f(1-aB(\theta(N_{B,S,\neq i}+N_{B,S,i}))-A_i)/Q-c'(N_{B,S,i})\right)+\frac{\partial}{\partial Z}\left(a\theta B'(\theta(N_{B,S,\neq i}+N_{B,S,i}))W\rho\pi f(1-aB(\theta(N_{B,S,\neq i}+N_{B,S,i}))-A_i)/Q-c'(N_{B,S,i})\right)=0$. Since we have assumed that second order conditions hold so $\frac{\partial}{\partial N_{B,S,i}}\left(a\theta B'(\theta(N_{B,S,\neq i}+N_{B,S,i}))W\rho\pi f(1-aB(\theta(N_{B,S,\neq i}+N_{B,S,i}))-A_i)/Q-c'(N_{B,S,i})\right)<0$, it follows that the sign of $\frac{\partial N_{B,S,i}}{\partial Z}$ is the same as the sign of $\frac{\partial}{\partial Z}\left(a\theta B'(\theta(N_{B,S,\neq i}+N_{B,S,i}))W\rho\pi f(1-aB(\theta(N_{B,S,\neq i}+N_{B,S,i}))-A_i)/Q-c'(N_{B,S,i})\right)$. It follows straightforwardly that $N_{B,S,i}$ is increasing with W, ρ and π. The derivative with respect to A_i is positive if and only if $f'(1-aB(\theta(N_{B,S,\neq i}+N_{B,S,i}))-A_i)<0$. The derivative with respect to $N_{B,S,\neq i}$ is negative, as long as $f'(1-aB(\theta(N_{B,S,\neq i}+N_{B,S,i}))-A_i)$ is not too negative. The derivative with respect to a is positive if and only if $af\left(1-aB\left(B(\theta N_{B,S})\right)-A_i\right)+B(\theta N_{B,S})af'\left(1-aB\left(B(\theta N_{B,S})\right)-A_i\right)>0$. The derivative with respect to θ is positive if and only if: $\frac{B\prime(\theta N_{B,S})+\theta N_{B,S}B"(\theta N_{B,S})}{aN_{B,S}B\prime(\theta N_{B,S})^2}>\frac{f"(1-aB(\theta N_{B,S})-A_i)}{f(1-aB(\theta N_{B,S})-A_i)}$.

Proof of Proposition 4: As $f'\left(-aB(\theta(N_{B,S,\neq i}+N_{B,S,i}))\right)>0$, the total derivative of $a\theta B'(\theta QN_{B,S,i})W\rho\pi f(-aB(\theta QN_{B,S,i}))/Q-c'(N_{B,S,i})$ with respect to $N_{B,S,i}$ must be positive. The sign of the impact of any exogenous variable Z on $N_{B,S,i}$ remains the same as the sign of its impact on $\theta B'(\theta QN_{B,S,i})W\rho\pi f(-aB(\theta QN_{B,S,i}))/Q$. The derivative of $N_{B,S,i}$ with respect to ρ, π, and W remains positive. The derivative of $B'(\theta QN_{B,S,i})f(-aB(\theta QN_{B,S,i}))/Q$ with respect to Q equals $\theta N_{B,S,i}B"(\theta QN_{B,S,i})f(-aB(\theta QN_{B,S,i}))/Q-a\theta N_{B,S,i}B'(\theta QN_{B,S,i})^2f'(-aB(\theta QN_{B,S,i}))/Q-B'(\theta QN_{B,S,i})f(-aB(\theta QN_{B,S,i}))/Q^2$. The slope of all three terms is negative, so the individual investment declines with Q. Aggregate investment declines with Q if and only if $QN_{B,S,i}$ is declining with Q, which requires that $-\frac{Q}{N_{B,S,i}}\frac{\partial N_{B,S,i}}{\partial Q}>1$. The value of $-\frac{Q}{N_{B,S,i}}\frac{\partial N_{B,S,i}}{\partial Q}$ is greater than one, if and only if $a\theta W\rho\pi\left(-\theta B"(\theta QN_{B,S,i})f(-aB(\theta QN_{B,S,i}))+a\theta B'(\theta QN_{B,S,i})^2f'(-aB(\theta QN_{B,S,i})+B'(\theta QN_{B,S,i})f(-aB(\theta QN_{B,S,i}))/QN_{B,S,i}\right)>\left(-a\theta^2W\rho\pi B"(\theta QN_{B,S,i})f(-aB(\theta QN_{B,S,i}))+a^2\theta^2W\rho\pi B'(\theta QN_{B,S,i})^2f'(-aB(\theta QN_{B,S,i}))\right)+c"(N_{B,S,i})$.

This will be true if and only if $c'(N_{B,S,i})>N_{B,S,i}c"(N_{B,S,i})$.

The derivative of $a\theta B'(\theta QN_{B,S,i})W\rho\pi f(-aB(\theta QN_{B,S,i}))/Q$ with respect to any variable that only works through B will be negative if and only if $\frac{\partial B\prime(\theta QN_{B,S,i})}{\partial Z}f(-aB(\theta QN_{B,S,i})) < aB'(\theta QN_{B,S,i})\frac{\partial B(\theta QN_{B,S,i})}{\partial Z}f'(-aB(\theta QN_{B,S,i}))$ If $f'(-aB(\theta QN_{B,S,i})$ is near zero, then this is simply the requirement that $\frac{\partial B\prime(\theta QN_{B,S,i})}{\partial Z} < 0$. The value of $\frac{\partial B\prime(\theta QN_{B,S,i})}{\partial N_{T,0}}$ equals positive terms times

$$\left(Log\left(\frac{p_0+\Delta}{p_0-\Delta}\right)p^* - (1-p^*)Log\left(\frac{1-p_0+\Delta}{1-p_0-\Delta}\right)\right)\left((p_0-\Delta)^{p^*N_{T,0}+\theta N_{B,S}}(1-p_0+\Delta)^{(1-p^*)N_{T,0}} - (p_0+\Delta)^{p^*N_{T,0}+\theta N_{B,S}}(1-p_0-\Delta)^{(1-p^*)N_{T,0}}\right)$$

The term $p^*Log\left(\frac{p_0+\Delta}{p_0-\Delta}\right) - (1-p^*)Log\left(\frac{1-p_0+\Delta}{1-p_0-\Delta}\right)$ is negative, when $p^* < \frac{Log\left(\frac{1-p_0+\Delta}{1-p_0-\Delta}\right)}{Log\left(\frac{1-p_0+\Delta}{1-p_0-\Delta}\right)+Log\left(\frac{p_0+\Delta}{p_0-\Delta}\right)} \approx \frac{p_0-\Delta}{1-2\Delta}$, which we assume. The term $\left((p_0-\Delta)^{p^*N_{T,0}+\theta N_{B,S}}(1-p_0+\Delta)^{(1-p^*)N_{T,0}} - (p_0+\Delta)^{p^*N_{T,0}+\theta N_{B,S}}(1-p_0-\Delta)^{(1-p^*)N_{T,0}}\right)$ is positive as long as $(p^*N_{T,0}+\theta N_{B,S})Log\left(\frac{p_0+\Delta}{p_0-\Delta}\right) < (1-p^*)N_{T,0}Log\left(\frac{1-p_0+\Delta}{1-p_0-\Delta}\right)$, which holds as long as $p^* < \frac{Log\left(\frac{1-p_0+\Delta}{1-p_0-\Delta}\right)-\frac{\theta N_{B,S}}{N_{T,0}}Log\left(\frac{p_0+\Delta}{p_0-\Delta}\right)}{Log\left(\frac{1-p_0+\Delta}{1-p_0-\Delta}\right)+Log\left(\frac{p_0+\Delta}{p_0-\Delta}\right)}$.

Proof of Proposition 5: With joint maximization, the first order condition is that $(\mu\theta_{MF} + (1-\mu)\theta_{FF})\delta W(H_F)(1-p_F a)\underline{t}_C = \alpha V'(N)$. Concavity implies that the number of children is increasing with α, p_F and a, and decreasing with $\underline{t}_C$, H_F, and δ. The number of children is increasing with μ if and only if $\theta_{MF} > \theta_{FF}$. Differentiation also gives that if $\theta_{FF} = \theta + \Delta$, and $\theta_{MF} = \theta - \Delta$, then the number of children is decreasing with θ and decreasing with Δ if and only if $\mu < .5$..

The derivative of the husband's welfare with respect to N equals $-\theta_{MF}\delta W(H_F)A_F\underline{t}_C + \alpha V'(N)$, which is positive if and only if $(1-\mu)(\theta_{FF} - \theta_{MF})\delta W(H_F)(1-p_F\alpha)\underline{t}_C > 0$, which is positive if and only if $\theta_{FF} > \theta_{MF}$. Let x denote $\delta W(H_F)(1-p_F\alpha)$ and then the husband's welfare is decreasing with x if and only if $(1-\mu)\left(\frac{\theta_{FF}}{\theta_{MF}} - 1\right)\frac{N\underline{t}_C}{1-N\underline{t}_C} > N\frac{-V''(N)}{V'(N)}$.

Proof of Proposition 6: We need only consider the case where the wife's ability level has generated a positive signal and the husband is considering destroying that signal. In that case, $N^- = V'^{-1}\left((\mu\theta_{MF} + (1-\mu)\theta_{FF})\delta W(H_F)\underline{t}_C/\alpha\right)$ denote the number of children chosen if the wife believes she is able and $N^+ = V'^{-1}\left(\left(\mu\theta_{MF} + (1-\mu)\theta_{FF}\left(1 - \frac{ap_F^0}{p_F^0+(1-p_F^0)(1-\pi+\pi\gamma g(k))}\right)\right)\delta W(H_F)\underline{t}_C/\alpha\right)$ denote the number of children chosen if the wife believes that she is not able. The concavity of V(.) ensures that $N^+ > N^-$.

If $1-\frac{ap_F^0}{p_F^0+(1-p_F^0)(1-\pi)} > \frac{\theta_{MF}}{\theta_{FF}}$, then the husband will benefit from increasing the number of children if the wife doesn't think that he has expended any effort on destroying the signal, and as such $\left(\alpha\big(V(N^+)-V(N^-)\big) > \theta_{MF}\delta W(H_F)\underline{t}_C(N^+-N^-)\right)$, at the point where k=0, and the condition $lim_{k\to 0}g'(k)=\infty$ guarantees that the husband will expend some effort destroying the wife's signal.

The first order condition for signal destruction is that $g'(k)\left(\alpha\big(V(N^+)-V(N^-)\big)-\theta_{MF}\delta W(H_F)\underline{t}_C(N^+-N^-)\right)=1$, and I left $h(N^+(k,x),N^-(x),x)$, denote $\alpha\big(V(N^+)-V(N^-)\big)-\theta_{MF}\delta W(H_F)\underline{t}_C(N^+-N^-)$, where x refers to any free parameter, so the first order condition can be rewritten $1=g'(k)h(N^+(k,x),N^-(x),x)$=1. The derivative of this with respect to any parameter x, delivers $\left(-g''(k)h-g'(k)h_{N^+}\frac{dN^+}{dk}\right)\frac{dk}{dx}=g'(k)\left(h_{N^+}\frac{dN^+}{dx}+h_{N^-}\frac{dN^-}{dx}+h_x\right)$, with $h_{N^+}\frac{dN^+}{dk}=\left(\alpha V'(N^+)-\theta_{MF}\delta W(H_F)\underline{t}_C\right)\frac{ap_F^0(1-p_F^0)\pi\gamma g\prime(k)(1-\mu)\theta_{FF}\delta W(H_F)\underline{t}_C}{\alpha V''(N^+)\left(p_F^0+(1-p_F^0)(1-\pi+\pi\gamma g(k))\right)^2}<0$. Thus the impact of any exogenous parameter on k has the same sign as $h_{N^+}\frac{dN^+}{dx}+h_{N^-}\frac{dN^-}{dx}+h_x$.

The parameters a, π, p_F^0 and γ, have no effect on N^- and no direct impact on h. Since $h_{N^+}>0$, there impact on k is the opposite of their impact on N^+. Since N^+ is rising with a, π and p_F^0 and falling with γ, the level of k is therefore increasing with a, π and p_F^0 and decreasing with γ.

The variables μ and θ_{FF} have no direct impact on h, but effect both N^- and N^+. The derivative $h_{N^+}\frac{dN^+}{dx}+h_{N^-}\frac{dN^-}{dx}$, equals

$$(1-\mu)\delta W(H_F)\underline{t}_C\left((\theta_{FF}-\theta_{MF})\left(\frac{dN^+}{dx}-\frac{dN^-}{dx}\right)-\frac{ap_F^0\theta_{FF}}{p_F^0+(1-p_F^0)(1-\pi+\pi\gamma g(k))}\frac{dN^+}{dx}\right)$$

Differentiation yields $\frac{dN^+}{d\mu}=\left(\theta_{MF}-\theta_{FF}\left(1-\frac{ap_F^0}{p_F^0+(1-p_F^0)(1-\pi+\pi\gamma g(k))}\right)\right)\frac{\delta W(H_F)\underline{t}_C}{\alpha V''(N^+)}$ and $\frac{dN^-}{d\mu}=(\theta_{MF}-\theta_{FF})\frac{\delta W(H_F)\underline{t}_C}{\alpha V''(N^-)}$, and if $V''(N^+)\approx V''(N^-)$,

$\frac{dN^+}{d\mu}-\frac{dN^-}{d\mu}\approx\frac{ap_F^0\theta_{FF}}{p_F^0+(1-p_F^0)(1-\pi+\pi\gamma g(k))}\frac{\delta W(H_F)\underline{t}_C}{\alpha V''(N^+)}$. Putting this together implies that:

$h_{N^+}\frac{dN^+}{d\mu}+h_{N^-}\frac{dN^-}{d\mu}$ equals $\frac{1}{V''(N^+)}\frac{(1-\mu)\left(\delta W(H_F)\underline{t}_C\right)^2 p_F^0\theta_{FF}}{p_F^0+(1-p_F^0)(1-\pi+\pi\gamma g(k))}$times $\left(2-\frac{ap_F^0}{p_F^0+(1-p_F^0)(1-\pi+\pi\gamma g(k))}\right)\theta_{FF}-2\theta_{MF}$, which is certainly negative if $1-\frac{ap_F^0}{p_F^0+(1-p_F^0)(1-\pi)}>\frac{\theta_{MF}}{\theta_{FF}}$.

Differentiation yields $\frac{dN^+}{d\theta_{FF}}=(1-\mu)\left(1-\frac{ap_F^0}{p_F^0+(1-p_F^0)(1-\pi+\pi\gamma g(k))}\right)\frac{\delta W(H_F)\underline{t}_C}{\alpha V''(N^+)}$ and $\frac{dN^-}{d\theta_{FF}}=(1-\mu)\frac{\delta W(H_F)\underline{t}_C}{\alpha V''(N^-)}$, and if $V''(N^+)\approx V''(N^-)$,

$\frac{dN^+}{d\mu}-\frac{dN^-}{d\mu}\approx-\frac{(1-\mu)p_F^0}{p_F^0+(1-p_F^0)(1-\pi+\pi\gamma g(k))}\frac{\delta W(H_F)\underline{t}_C}{V''(N^+)}$. Putting this together implies that $h_{N^+}\frac{dN^+}{d\mu}+h_{N^-}\frac{dN^-}{d\mu}$ equals $\frac{1}{V''(N^+)}\frac{\left((1-\mu)\delta W(H_F)\underline{t}_C\right)^2 p_F^0}{p_F^0+(1-p_F^0)(1-\pi+\pi\gamma g(k))}$ times $\left(\theta_{MF}-\theta_{FF}\left(2-\frac{ap_F^0}{p_F^0+(1-p_F^0)(1-\pi+\pi\gamma g(k))}\right)\right)$, which is positive as long as $1-\frac{ap_F^0}{p_F^0+(1-p_F^0)(1-\pi)}>\frac{\theta_{MF}}{\theta_{FF}}$.

$\alpha\left(V(N^+)-V(N^-)\right)-\theta_{MF}\delta W(H_F)\underline{t}_C(N^+-N^-)$,

I know let z denote $\delta W(H_F)\underline{t}_C$, and find that $h_{N^+}\frac{dN^+}{dz}+h_{N^-}\frac{dN^-}{dz}+h_z$

$$(1-\mu)z\left((\theta_{FF}-\theta_{MF})\left(\frac{dN^+}{dz}-\frac{dN^-}{dz}\right)-\frac{ap_F^0\theta_{FF}}{p_F^0+(1-p_F^0)(1-\pi+\pi\gamma g(k))}\frac{dN^+}{dz}\right)$$
$$-\theta_{MF}(N^+-N^-)$$

Differentiation yields $\frac{dN^+}{dz}=\frac{\mu\theta_{MF}+(1-\mu)\theta_{FF}\left(1-\frac{ap_F^0}{p_F^0+(1-p_F^0)(1-\pi+\pi\gamma g(k))}\right)}{\alpha V''(N^+)}$ and $\frac{dN^-}{dz}=\frac{\mu\theta_{MF}+(1-\mu)\theta_{FF}}{\alpha V''(N^+)}$, and if $V''(N^+)\approx V''(N^-)$, $\frac{dN^+}{dz}-\frac{dN^-}{dz}\approx-\frac{ap_F^0(1-\mu)\theta_{FF}}{\alpha V''(N^+)\left(p_F^0+(1-p_F^0)(1-\pi+\pi\gamma g(k))\right)}$. If third order terms are small, then we can use the Taylor series approximation that $\alpha V'(N^-)+\alpha V''(N^-)(N^+-N^-)\approx\alpha V'(N^+)$, which implies that $(N^+-N^-)\approx\frac{-(1-\mu)\theta_{FF}ap_F^0 z}{\alpha V''(N^-)(p_F^0+(1-p_F^0)(1-\pi+\pi\gamma g(k)))}$, so that implies that $h_{N^+}\frac{dN^+}{dz}+h_{N^-}\frac{dN^-}{dz}+h_z\frac{z}{\alpha V''(N^+)}\left(\frac{(1-\mu)\theta_{FF}ap_F^0}{p_F^0+(1-p_F^0)(1-\pi+\pi\gamma g(k))}\right)^2<0$. Increases in women's education, or the time cost of children, or reductions in discrimination against women in the labor market, will lead to less effort reducing women's assessment of their own ability.

Proof of Proposition 7: The first order condition for parental investment in a daughter's education is that

$$1 = \alpha_1\big(\delta W'(H_F)A_F(1 - N\underline{t}_C\ )\big) + \big((\alpha_1\alpha + \alpha_2)V'(N) - \alpha_1\delta W(H_F)A_F\underline{t}_C\big)\frac{\partial \mathrm{N}}{\partial H_F}$$

The child-bearing choices during the next generation set $\delta W(H_F)A_F\underline{t}_C = \alpha V'(N)$, so $\frac{\partial \mathrm{N}}{\partial H_F}$ equals $\frac{\delta W\prime(H_F)A_F\underline{t}_C}{\alpha V"(N)}$ and the first order condition for parental investment can be written as

$1 = W'(H_F)\delta A_F\alpha_1\left(1 - \left(1 - \frac{\alpha_2 V\prime(N)}{\alpha_1\alpha N V"(N)}\right)N\underline{t}_C\right)$, which can be written as
$1 = W'(H_F(x))g(N(H_F(x),x),x)$, where x represents any of the other exogenous variables.

If $lim_{x\to 0}W'(x) = \infty$ and if $\frac{V"(N)}{V\prime(N)} = \frac{\sigma-1}{N}$, then the condition for positive investment in human capital for daughters is that $1 > \left(1 + \frac{\alpha_2}{\alpha_1\alpha}\right)\mathrm{N}\underline{t}_C$.

Differentiating this equation totally with respect to any "x" produces $-\Big(W"(H_F(x))g(N,x) + W'(H_F(x))\frac{\partial \mathrm{g}}{\partial N}\frac{\partial \mathrm{N}}{\partial H_F}\Big)\frac{\partial H_F}{\partial \mathrm{x}} = W'(H_F(x))\left(\frac{\partial \mathrm{g}}{\partial N}\frac{\partial \mathrm{N}}{\partial x} + \frac{\partial \mathrm{g}}{\partial x}\right)$, I assume that second order conditions hold, which requires that $-\frac{\alpha V"(N)W"(H_F)}{\delta A_F\underline{t}_C\left(W\prime(H_F)\right)^3} < \frac{\partial \mathrm{g}}{\partial N} = \delta A_F\underline{t}_C\left(\frac{\alpha_2}{\alpha} + \frac{\alpha_2}{\alpha}\frac{V\prime\prime\prime(N)V\prime(N)}{(V"(N))^2} - \alpha_1\right)$ or $\alpha_1 - \frac{\alpha V"(N)W"(H_F)}{\left(\delta A_F\underline{t}_C\right)^2\left(W\prime(H_F)\right)^3} < \frac{\alpha_2}{\alpha}\left(1 + \frac{V\prime\prime\prime(N)V\prime(N)}{(V"(N))^2}\right)$.

With that assumption, the sign of $\frac{\partial H_F}{\partial \mathrm{x}}$ depends on the sign of $\frac{\partial \mathrm{g}}{\partial N}\frac{\partial \mathrm{N}}{\partial x} + \frac{\partial \mathrm{g}}{\partial x}$. The terms α_1 and α_2 do not directly impact N, and since $\frac{\partial \mathrm{g}}{\partial \alpha_1} = A_F\big(1 - N\underline{t}_C\ \big) > 0$ and $\frac{\partial \mathrm{g}}{\partial \alpha_1} = \frac{V'(N)\delta A_F\underline{t}_C}{\alpha V"(N)} < 0$, we know that investment in the daughter's human capital is rising in direct altruism for the direct and declining with altruism towards the number of grandchildren.

The term δA_F can be labeled y, since these two terms always enter together. In that case, $\frac{\partial \mathrm{g}}{\partial N}\frac{\partial \mathrm{N}}{\partial y} + \frac{\partial \mathrm{g}}{\partial y}$, which is positive as long as $\frac{1}{\delta A_F W'(H_F)} + \frac{\delta W(H_F)}{\alpha V"(N)}\frac{\partial \mathrm{g}}{\partial N} > 0$, or $\frac{-\alpha V"(N)}{\delta^2 A_F W\prime(H_F)W(H_F)} > \frac{\partial \mathrm{g}}{\partial N}$. At the maximum $\frac{1}{(1-N\underline{t}_C\ )\delta A_F W'(H_F)} - \frac{\alpha_2 V\prime(N)\underline{t}_C}{(1-N\underline{t}_C\ )\alpha V"(N)} = \alpha_1$ so $\frac{\partial \mathrm{g}}{\partial N}$ can also be written as

$$\alpha_1 - \frac{\alpha V"(N)}{\delta^3 A_F^2\underline{t}_C W'(H_F)W(H_F)} > \frac{\alpha_2}{\alpha}\left(1 + \frac{V'''(N)V'(N)}{(V"(N))^2}\right)$$

Which is always positive if α_2 is sufficiently small.

$$\delta A_F\underline{t}_C\left(\frac{\alpha_2}{\alpha} + \frac{\alpha_2}{\alpha}\frac{V'''(N)V'(N)}{(V"(N))^2} - \alpha_1\right)$$

$$\delta A_F\underline{t}_C\left(\frac{\alpha_2}{\alpha} + \frac{\alpha_2}{\alpha}\frac{V'''(N)V'(N)}{(V"(N))^2} - \frac{1}{(1 - N\underline{t}_C\ )\delta A_F W'(H_F)} + \frac{\alpha_2 V'(N)\underline{t}_C}{(1 - N\underline{t}_C\ )\alpha V"(N)}\right)$$

$\frac{1}{\delta A_F W'(H_F)}\left(\frac{-\alpha V''(N)}{\delta^2 A_F \underline{t}_C W(H_F)}+\frac{1}{1-N\underline{t}_C}\right)>\frac{\alpha_2}{\alpha}\left(1+\frac{V'''(N)V'(N)}{(V''(N))^2}+\frac{V'(N)\underline{t}_C}{(1-N\underline{t}_C\ )V''(N)}\right).$

Proof of Lemma 1: Define skilled parental welfare at H given a deviation that is perceived as having come from an unskilled child's parent as $G(H, N_U(H), x)$. For H sufficiently close to H_F^{Skill}, $G(H, N_U(H), x)$ is greater than the payoff to choosing H_F^{Skill} and for H sufficiently close to 0, the parents of skilled parents strictly prefer choosing H_F^{Skill}. By continuity, there must exist at least one value of H so such that $G(H, N_U(H), x)$ equals $G\left(H_F^{Skill}, N_S\left(H_F^{Skill}\right), x\right)$ the welfare of parents of skilled children when they choose H_F^{Skill}.

The derivatives $\frac{\partial G}{\partial H}+\frac{\partial G}{\partial N_U}\frac{\partial N_U}{\partial H}$ equal $\alpha_1\left(\delta W'(H)(1-N_U(H)\underline{t}_C\ )\right)+\alpha_2 V'(N_U)\frac{\delta W'(H)(1-a)\underline{t}_C}{\alpha V''(N_U)}$.

The derivative of this expression with respect to H is $\delta(1-a)$ times

$W''(H)\left(\frac{\alpha_1}{1-a}\left((1-N_U\underline{t}_C\ )\right)+\frac{\alpha_2 V'(N_U)\underline{t}_C}{\alpha V''(N_U)}\right)+\left(\frac{\alpha_2}{\alpha}-\frac{\alpha_1}{1-a}-\frac{\alpha_2 V'(N_U)V''(H)}{\alpha(V''(N_U))^2}\right)\frac{\delta\left(\underline{t}_C W'(H)\right)^2(1-a)}{\alpha V''(N_U))}$ which is equal to $W''(H)\left(\alpha_1\left((1-N_U\underline{t}_C\ )\right)+\frac{\alpha_2 V'(N_U)\underline{t}_C}{\alpha V''(N_U)}\right)+\left(\frac{\alpha_2}{\alpha}-\alpha_1-\frac{\alpha_2 V'(N_U)V''(H)}{\alpha(V''(N_U))^2}\right)\frac{\delta\left(\underline{t}_C W'(H)\right)^2(1-a)}{\alpha V''(N_U))}$ plus $W''(H)\frac{\alpha_1 a}{1-a}\left((1-N_U\underline{t}_C\ )\right)-\frac{a\alpha_1}{1-a}\frac{\delta\left(\underline{t}_C W'(H)\right)^2(1-a)}{\alpha V''(N_U))}$. The first expression is negative by assumption (we assumed the maximization problem was concave for skilled and unskilled parents). The second expression is negative as long as $W''(H)\left((1-N_U\underline{t}_C\ )\right)<\frac{\delta\left(\underline{t}_C W'(H)\right)^2(1-a)}{\alpha V''(N_U))}$, which also must be true since $W''(H)\left(\alpha_1\left((1-N_U\underline{t}_C\ )\right)+\frac{\alpha_2 V'(N_U)\underline{t}_C}{\alpha V''(N_U)}\right)+\left(\frac{\alpha_2}{\alpha}-\alpha_1-\frac{\alpha_2 V'(N_U)V''(H)}{\alpha(V''(N_U))^2}\right)\frac{\delta\left(\underline{t}_C W'H\right)^2(1-a)}{\alpha V''(N_U))}$ must be negative at $\alpha_2=0$.

Since the function $G(H, N_U(H), x)$ is concave and begins below $G\left(H_F^{Skill}, N_S\left(H_F^{Skill}\right), x\right)$ and ends above $G\left(H_F^{Skill}, N_S\left(H_F^{Skill}\right), x\right)$ for H close to H_F^{Skill}, there can exist either one or two crossing points. There cannot exist more than two crossing points, because the slope of $G(H, N_U(H), x)$ can only change sign once. If there exists only one crossing point then the slope of $G(H, N_U(H), x)$ must be positive at the crossing point, since otherwise $G(H, N_U(H), x)$ would be below $G\left(H_F^{Skill}, N_S\left(H_F^{Skill}\right), x\right)$ for all lower values of H and we know that is false. If $G(H, N_U(H), x)$ crosses twice, then the slope must be positive at the lower value of H, which defines $\widehat{H}$, since the function is concave.

The expression that defines $\widehat{H}$ can be written as $G\left(H_F^{Skill}, N_S\left(H_F^{Skill}\right), x\right)=G\left(\widehat{H}, N_U\left(\widehat{H}\right), x\right)$, and the derivative of this with respect to any parameter x satisfies: $\left(\frac{\partial G}{\partial H_F^{Skill}}+\frac{\partial G}{\partial N_S}\frac{\partial N_S}{\partial H_F^{Skill}}\right)\frac{\partial H_F^{Skill}}{\partial x}+$

$\frac{\partial G(H_F^{Skill}, N_S(H_F^{Skill}), x)}{\partial x} = \left(\frac{\partial G}{\partial \hat{H}} + \frac{\partial G}{\partial \hat{H}} \frac{\partial N_U}{\partial H_F^{Skill}}\right) \frac{\partial \hat{H}}{\partial x} + \frac{\partial G(\hat{H}, N_U(\hat{H}), x)}{\partial x}$. Since $G(H_F^{Skill}, N_S(H_F^{Skill}), x)$ is maximized over H_F^{Skill}, then envelope theorem applies, and $\frac{\partial G}{\partial H_F^{Skill}} + \frac{\partial G}{\partial N_S} \frac{\partial N_S}{\partial H_F^{Skill}} = 0$. Moreover, the derivatives $\frac{\partial G}{\partial \hat{H}} + \frac{\partial G}{\partial N_U} \frac{\partial N_U}{\partial \hat{H}}$ is positive since the slope of $G(\hat{H}, N_U(\hat{H}), x)$ is positive at the crossing point. That means that the sign of $\frac{\partial \hat{H}}{\partial x}$ equals the sign of $\frac{\partial G(H_F^{Skill}, N_S(H_F^{Skill}), x)}{\partial x} - \frac{\partial G(\hat{H}, N_U(\hat{H}), x)}{\partial x}$.

For x=δ, the value of $\frac{\partial G(H_F^{Skill}, N_S(H_F^{Skill}), x)}{\partial x} - \frac{\partial G(\hat{H}, N_U(\hat{H}), x)}{\partial x}$ equals α_1 times $W(H_F^{Skill})(1 - N_S(H_F^{Skill})\underline{t}_C) - W(\hat{H})(1 - N_U(\hat{H})\underline{t}_C)$, which is strictly positive since $W(H_F^{Skill}) > W(\hat{H})$ and $N_S(H_F^{Skill}) < N_U(\hat{H})$.

For x=$\underline{t}_C$, the value of $\frac{\partial G(H_F^{Skill}, N_S(H_F^{Skill}), x)}{\partial x} - \frac{\partial G(\hat{H}, N_U(\hat{H}), x)}{\partial x}$ equals $V\left(N_S(H_F^{Skill})\right) - V\left(N_U(\hat{H})\right)$ is strictly negative

For x=α_2, the value of $\frac{\partial G(H_F^{Skill}, N_S(H_F^{Skill}), x)}{\partial x} - \frac{\partial G(\hat{H}, N_U(\hat{H}), x)}{\partial x}$ equals $V\left(N_S(H_F^{Skill})\right) - V\left(N_U(\hat{H})\right)$ is strictly negative.

For x=α_1, the value of $\frac{\partial G(H_F^{Skill}, N_S(H_F^{Skill}), x)}{\partial x} - \frac{\partial G(\hat{H}, N_U(\hat{H}), x)}{\partial x}$ equals $\delta W(H_F^{Skill})(1 - N_S(H_F^{Skill})\underline{t}_C) + \alpha V\left(N_S(H_F^{Skill})\right) - \delta W(\hat{H})(1 - N_U(\hat{H})\underline{t}_C) - \alpha V(N_U(\hat{H}))$, since the derivative of $\delta W(H)(1 - N_S(H)\underline{t}_C) + \alpha V(N_S(H))$ with respect to H is $\delta W'(H)(1 - N_S(H)\underline{t}_C)$ and $N_U(\hat{H}) > N_S(\hat{H})$.

Proof of Proposition 8:

If $H_F^{Unskill} < \hat{H}$, then there exists a separating equilibrium entails the unskilled choosing $H_F^{Unskill}$ and the skilled choose H_F^{Skill}. These are equilibria, and they are stable, because neither group wants to imitate each other. The parents of less skilled daughters have no incentive to deviate since they are already receiving their best possible outcome. As such, any deviation can only come from the skilled. Since D1 requires then that children believe that any deviation can only come from the skilled, there is no reason for the skilled to deviate either.

Are there other separating equilibria when $H_F^{Unskill} < \hat{H}$. The parents of the skilled are always better off at H_F^{Skill}, then at any other separating human capital level, so there can be no equilibrium where the parents of the skilled choose a different level of H. That fact is true even without the equilibrium refinement.

Consider any other choice o H_F' where the unskilled will pool, and consider a deviation to $H_F^{Unskill}$. There is no set of beliefs by children, or fertility response, that could induce the parent of the skilled to make that deviation, and so only the unskilled will make it and this means that the only reasonable beliefs by children is that they are unskilled given a choice of $H_F^{Unskill}$. That response means that the unskilled will deviate to $H_F^{Unskill}$.

It is similarly impossible to generate a pooling or semi-pooling equilibrium, in this range of parameter values since the unskilled can always deviate to $H_F^{Unskill}$. In this pooling equilibrium, the parents of the unskilled are always doing worse than in the separating equilibrium (since it is their best outcome) and the parents of the skilled are always doing better (since they always have the option to separate). The parents of the skilled will therefore happily deviate if the fertility response is $N_U\left(H_F^{Unskill}\right)$ while the parents of the unskilled will require a higher fertility level to deviate. This means that children will infer that such a deviation can only come from the less skilled, and this inference will cause the parents

If $H_F^{Unskill} > \widehat{H} > \underline{H}_F$ then the only separating equilibrium will again require that the parents of the skilled invest H_F^{Skill} as discussed above. There does not exist a separating equilibrium where the parents of the less skilled choose a value of H greater than $\widehat{H}$ because the parents of the more skilled would choose to imitate them and that does not require our refinement. Since the welfare problem of the parents of the unskilled is concave, they would strictly prefer a higher level of human capital investment, given that their daughters think they are unskilled. There does not exist a separating equilibrium where the parents of the less skilled choose a lower level of H, because if they deviate to $\widehat{H} - \varepsilon$, then children must realize that if they choose $N_U\left(\widehat{H} - \varepsilon\right)$, the parents of the unskilled would strictly prefer this outcome. But if they choose $N_U\left(\widehat{H} - \varepsilon\right)$, the parents of the skilled would strictly prefer to remain investing $N_U\left(H_F^{Unskill}\right)$. As a result, they must infer that the deviation came from the less skilled and this means that the deviation is thought to come from parents of the unskilled, and as a result those parents will deviate.

A separating equilibrium does exist where the parents of the unskilled choose $\widehat{H}$ and the parents of the skilled choose H_F^{Skill}. The parents of the skilled will not choose to deviate to $\widehat{H}$, since they are getting equal welfare by choosing H_F^{Skill}. The parents of the unskilled will not choose to deviate to any lower level of H, because their welfare is locally increasing in H, assuming that children think that they are unskilled.

What are the beliefs of children who receive a level of investment $H > \widehat{H}$? Consider a candidate deviation H*>$\widehat{H}$, and let N_U^* denote the fertility level that makes the parents of the unskilled indifferent with deviating to $H*$ and N_S^* denote the fertility level that makes the parents of the skilled indifferent with deviating to $H*$. These terms are defined by

$$\alpha_1\left((1-a)\delta W(H^*)\left(1-N_U^*\underline{t}_C\right)+\alpha V(N_U^*)\right)+\alpha_2 V(N_U^*)-H^*$$
$$=\alpha_1\left((1-a)\delta W\left(\widehat{H}\right)(1-N_U(\widehat{H})\underline{t}_C)+\alpha V(N_U(\widehat{H}))\right)+\alpha_2 V\left(N_U(\widehat{H})\right)-\widehat{H}$$

and

$$\alpha_1\left((1-a)\delta W(H^*)\left(1-N_S^*\underline{t}_C\right)+\alpha V(N_S^*)\right)+\alpha_2 V(N_S^*)-H^*+\alpha_1 a\left(\delta W(H^*)\left(1-N_S^*\underline{t}_C\right)-\delta W\left(\widehat{H}\right)\left(1-N_U(\widehat{H})\underline{t}_C\right)\right)=\alpha_1\left((1-a)\delta W\left(\widehat{H}\right)\left(1-N_U(\widehat{H})\underline{t}_C\right)+\alpha V(N_U(\widehat{H}))\right)+\alpha_2 V\left(N_U(\widehat{H})\right)-\widehat{H}.$$

Since $N_S^* < N_U(\widehat{H})$, it follows that $\alpha_1 a\left(\delta W(H^*)\left(1-N_S^*\underline{t}_C\right)-\delta W\left(\widehat{H}\right)\left(1-N_U(\widehat{H})\underline{t}_C\right)\right)>0$. Since $\alpha_1\left((1-a)\delta W(H^*)\left(1-N\underline{t}_C\right)+\alpha V(N)\right)+\alpha_2 V(N)-H^*$ is increasing with N (as long as fertility is rationally chosen by children), then this implies that $N_S^* < N_U^*$, and this means that the any deviation is thought to come from the parents of skilled parents.

To prove that no parents of unskilled children would want to deviate given those beliefs, let $\widetilde{H}$ denote the value of H that maximizes $\alpha_1\left((1-a)\delta W\left(\widetilde{H}\right)\left(1-N_S(\widetilde{H})\underline{t}_C\right)+\alpha V(N_S(\widetilde{H}))\right)+\alpha_2 V\left(N_U(\widetilde{H})\right)-\widetilde{H}$. Then the condition for deviation is that

$$\alpha_1\left((1-a)\delta W\left(\widetilde{H}\right)\left(1-N_S(\widetilde{H})\underline{t}_C\right)+\alpha V(N_S(\widetilde{H}))\right)+\alpha_2 V\left(N_U(\widetilde{H})\right)-\widetilde{H}>\alpha_1\left((1-a)\delta W\left(\widehat{H}\right)\left(1-N_U(\widehat{H})\underline{t}_C\right)+\alpha V(N_U(\widehat{H}))\right)+\alpha_2 V\left(N_U(\widehat{H})\right)-\widehat{H}.$$

or

$$\alpha_1\left(\delta W\left(\widetilde{H}\right)\left(1-N_S(\widetilde{H})\underline{t}_C\right)+\alpha V(N_S(\widetilde{H}))\right)+\alpha_2 V\left(N_U(\widetilde{H})\right)-\widetilde{H}-\ \alpha_1 a\left(\delta W\left(\widetilde{H}\right)\left(1-N_S(\widetilde{H})\underline{t}_C\right)-\delta W\left(\widehat{H}\right)\left(1-N_U(\widehat{H})\underline{t}_C\right)\right)>$$
$$\alpha_1\left(\delta W\left(H_F^{Skill}\right)\left(1-N_S\left(H_F^{Skill}\right)\underline{t}_C\right)+\alpha V\left(N_S\left(H_F^{Skill}\right)\right)\right)+\alpha_2 V\left(N_S\left(H_F^{Skill}\right)\right)-H_F^{Skill}.$$

But that is impossible since

$$\alpha_1\left(\delta W\left(\widetilde{H}\right)\left(1-N_S(\widetilde{H})\underline{t}_C\right)+\alpha V(N_S(\widetilde{H}))\right)+\alpha_2 V\left(N_U(\widetilde{H})\right)-\widetilde{H} < \alpha_1\left(\delta W\left(H_F^{Skill}\right)\left(1-N_S\left(H_F^{Skill}\right)\underline{t}_C\right)+\alpha V\left(N_S\left(H_F^{Skill}\right)\right)\right)+\alpha_2 V\left(N_S\left(H_F^{Skill}\right)\right)-H_F^{Skill}$$

because H_F^{Skill} maximizes $\alpha_1\left(\delta W(H)(1-N_S(H)\underline{t}_C)+\alpha V(N_S(H))\right)+\alpha_2 V(N_U(H)-H$ and $\left(W\left(\widetilde{H}\right)\left(1-N_S(\widetilde{H})\underline{t}_C\right)-W\left(\widehat{H}\right)\left(1-N_U(\widehat{H})\underline{t}_C\right)\right)>0$, as $N_S\left(\widetilde{H}\right)>N_U(\widehat{H})$.

As such this equilibrium exists and is stable.

To rule out any pooling or semi-pooling equilibrium, let N_{Pool} denote the fertility level at this pooled or semi-pooled equilibrium, where H equals H_{Pool} and let N_U^* denote the fertility level that makes the parents of the unskilled indifferent with deviating to $H_{Pool}-\varepsilon$ and N_S^* denote the fertility level that makes the parents of the skilled indifferent with deviating to $H_{Pool}-\varepsilon$. Note further that in a pooling, or semi-pooling equilibrium, the welfare for the skilled must be greater than if they choose H_F^{Skill} since they always have that option. These terms are defined by

$$\alpha_1\left((1-a)\delta W(H_{Pool}-\varepsilon)(1-N_U^*\underline{t}_C) + \alpha V(N_U^*)\right) + \alpha_2 V(N_U^*) - H_{Pool} + \varepsilon$$
$$= \alpha_1\big((1-a)\delta W(H_{Pool})(1-N_{Pool}\underline{t}_C) + \alpha V(N_{Pool})\big) + \alpha_2 V(N_{Pool}) - H_{Pool}$$

And

$$\alpha_1\left(\delta W(H_{Pool}-\varepsilon)(1-N_S^*\underline{t}_C) + \alpha V(N_S^*)\right) + \alpha_2 V(N_S^*) - H_{Pool} + \varepsilon$$
$$= \alpha_1\big(\delta W(H_{Pool})(1-N_{Pool}\underline{t}_C) + \alpha V(N_{Pool})\big) + \alpha_2 V(N_{Pool}) - H_{Pool}$$

The first equality can be written as

$$\alpha_1\left(\delta W(H_{Pool}-\varepsilon)(1-N_U^*\underline{t}_C) + \alpha V(N_U^*)\right) + \alpha_2 V(N_U^*) - H_{Pool} + \varepsilon$$
$$= \alpha_1 a\delta\big(W(H_{Pool}-\varepsilon)(1-N_U^*\underline{t}_C) - W(H_{Pool})(1-N_{Pool}\underline{t}_C)\big)$$
$$+ \alpha_1\big(\delta W(H_{Pool})(1-N_{Pool}\underline{t}_C) + \alpha V(N_{Pool})\big) + \alpha_2 V(N_{Pool}) - H_{Pool}$$

The expression $\alpha_1 a\delta\big(W(H_{Pool}-\varepsilon)(1-N_U^*\underline{t}_C) - W(H_{Pool})(1-N_{Pool}\underline{t}_C)\big)$ must be negative since $W(H_{Pool}-\varepsilon) < W(H_{Pool})$ and $N_U^* > N_{Pool}$ (since higher fertility must offset the lower investment level). Since $\alpha_1\left(\delta W(H_{Pool}-\varepsilon)(1-N\underline{t}_C) + \alpha V(N)\right) + \alpha_2 V(N) - H_{Pool} + \varepsilon$ is increasing with N (given rational behavior by the children), it follows that $N_U^* < N_S^*$ and the deviation must be thought to come from parents of the less skilled. As a result, children must infer that the deviation comes from them, and the parents of the less skilled will always strictly prefer deviation downward for a sufficiently small value of ε.

When $\widehat{H} < \underline{H}_F$, then the equilibrium involves pooling at this point, and some parents of the skilled may remain at H_F^{Skill}. We can rule out pure separating, since the parents of the more skilled will always choose H_F^{Skill} and will strictly prefer imitating the less skilled at any feasible level of H. We can rule out any pooling at a higher level of H, by the argument made immediately above, since the parents of the less skilled will always choose to deviate downwards. That leaves us with pure pooling, or semi-pooling. Pure pooling can only exist if the welfare of the parents of the skilled at the pooling point is greater than the welfare they would have by choosing H_F^{Skill} and being known to be more skilled.

At the pooling point, any deviation upwards will be thought to come from the more skilled, by a variant of the argument made above, and again, following roughly the algebra above, the less skilled will not choose to deviate. Let $p_{Mix} \leq p_F$, denote the probability of being skilled condition upon receiving investment level $\underline{H}_F$ and let $N_{Mix}(\underline{H}_F)$ denote the optimal fertility response given that belief. Then consider any deviation to a higher level of H denoted H'. Let N_U^* denote the fertility level that makes the parents of the unskilled indifferent with deviating to H' and N_S^* denote the fertility level that makes the parents of the skilled indifferent with deviating to H'.

These terms are defined by

$$\alpha_1\left((1-a)\delta W(H')\left(1-N_U^*\underline{t}_C\right)+\alpha V(N_U^*)\right)+\alpha_2 V(N_U^*)-H'$$
$$=\alpha_1\left((1-a)\delta W\left(\underline{H}_F\right)(1-N_{Mix}(\underline{H}_F)\underline{t}_C)+\alpha V(N_{Mix}(\underline{H}_F))\right)+\alpha_2 V\left(N_{Mix}(\underline{H}_F)\right)-\underline{H}_F$$

and

$$\alpha_1\left((\delta W(H')\left(1-N_S^*\underline{t}_C\right)+\alpha V(N_S^*)\right)+\alpha_2 V(N_S^*)-H'$$
$$=\alpha_1\left(\delta W\left(\underline{H}_F\right)(1-N_{Mix}(\underline{H}_F)\underline{t}_C)+\alpha V(N_{Mix}(\underline{H}_F))\right)+\alpha_2 V\left(N_{Mix}(\underline{H}_F)\right)-\underline{H}_F$$

This second equality can be rewritten as:

$$\alpha_1\left((1-a)(\delta W(H')\left(1-N_S^*\underline{t}_C\right)+\alpha V(N_S^*)\right)+\alpha_2 V(N_S^*)-H'$$
$$+\alpha_1\delta\left(W(H')\left(1-N_S^*\underline{t}_C\right)-W\left(\underline{H}_F\right)(1-N_{Mix}(\underline{H}_F)\underline{t}_C\right)$$
$$=\alpha_1\left((1-a)\delta W\left(\underline{H}_F\right)(1-N_{Mix}(\underline{H}_F)\underline{t}_C)+\alpha V(N_{Mix}(\underline{H}_F))\right)+\alpha_2 V\left(N_{Mix}(\underline{H}_F)\right)-\underline{H}_F$$

As $N_S^* < N_{Mix}(\underline{H}_F)$, since a lower fertility level must offset for higher human capital investment for parents of the more skilled, it follows that $N_S^* < N_U^*$ and deviations are more likely to come from the parents of the skilled. As the parents of the skilled will not deviate if their children then know that they are skilled, since their deviation would always yield a lower welfare level than choosing H_F^{Skill}, the parents of the less skilled will also choose not to deviate.

The value of p_{Mix} ensures a fertility level (chosen by children following the first order condition $\delta W\left(\underline{H}_F\right)(1-p_{Mix}a)\underline{t}_C=\alpha V'(N_{Mix}\left(\underline{H}_F\right)))$ so that

$$\alpha_1\left(\delta W\left(\underline{H}_F\right)(1-N_{Mix}\underline{t}_C)+\alpha V(N_{Mix})\right)+\alpha_2 V(N_{Mix})-\underline{H}_F$$
$$=\alpha_1\left(\delta W\left(H_F^{Skill}\right)\left(1-N_S\left(H_F^{Skill}\right)\underline{t}_C\right)+\alpha V\left(N_S\left(H_F^{Skill}\right)\right)\right)+\alpha_2 V\left(N_S\left(H_F^{Skill}\right)\right)-H_F^{Skill}$$

This equality can be written as The expression that defines $\widehat{H}$ can be written as $G\left(H_F^{Skill},N_S\left(H_F^{Skill}\right),x\right)=G\left(\underline{H}_F,N_{Mix}\left(\underline{H}_F,p_{Mix}(x)\right)\right),x)$. Using $\frac{\partial G}{\partial H_F^{Skill}}+\frac{\partial G}{\partial N_S}\frac{\partial N_S}{\partial H_F^{Skill}}=0$, differentiating this expression totally with respect to x yields:

$\frac{\partial G\left(H_F^{Skill},N_S\left(H_F^{Skill}\right),x\right)}{\partial x}-\frac{\partial G(\widehat{H},N_U(\widehat{H}),x)}{\partial x}-\frac{\partial G}{\partial N_{Mix}}\frac{\partial N_{Mix}}{\partial x}=\frac{\partial G}{\partial N_{Mix}}\frac{\partial N_{Mix}}{\partial p_{Mix}}\frac{\partial p_{Mix}}{\partial x}$. We know that $\frac{\partial G}{\partial N_{Mix}}\frac{\partial N_{Mix}}{\partial p_{Mix}}>0$, so the sign of $\frac{\partial p_{Mix}}{\partial x}$ is the same as the sign of $\frac{\partial G\left(H_F^{Skill},N_S\left(H_F^{Skill}\right),x\right)}{\partial x}-\frac{\partial G(\widehat{H},N_U(\widehat{H}),x)}{\partial x}-\frac{\partial G}{\partial N_{Mix}}\frac{\partial N_{Mix}}{\partial x}$. We first consider α_1 and α_2 which do not directly impact fertility.

When x=α_1 then $\frac{\partial G(H_F^{Skill}, N_S(H_F^{Skill}), x)}{\partial x} - \frac{\partial G(\hat{H}, N_U(\hat{H}), x)}{\partial x}$ equals $\left(\delta W(H_F^{Skill})(1 - N_S(H_F^{Skill})\underline{t}_C) + \alpha V\left(N_S(H_F^{Skill})\right)\right) - \left(\delta W(\underline{H}_F)(1 - N_{Mix}\underline{t}_C) + \alpha V(N_{Mix})\right)$ which must be positive since $\delta W(H)(1 - N_S(H)\underline{t}_C) + \alpha V(N_S(H))$ is strictly increasing with H and $\delta W(\underline{H}_F)(1 - N_S(\underline{H}_F)\underline{t}_C) + \alpha V\left(N_S(\underline{H}_F)\right) > \left(\delta W(\underline{H}_F)(1 - N_{Mix}\underline{t}_C) + \alpha V(N_{Mix})\right)$. Hence the number of parents of more skilled children choosing $\underline{H}_F$ must decline with α_1.

When x=α_2 then $\frac{\partial G(H_F^{Skill}, N_S(H_F^{Skill}), x)}{\partial x} - \frac{\partial G(\hat{H}, N_U(\hat{H}), x)}{\partial x}$ equals $V\left(N_S(H_F^{Skill})\right) - V(N_{Mix})$ which must be negative. Hence the number of parents of more skilled children choosing $\underline{H}_F$ must decline with α_2.

When x=δ, $\frac{\partial G(H_F^{Skill}, N_S(H_F^{Skill}), x)}{\partial x} - \frac{\partial G(\hat{H}, N_U(\hat{H}), x)}{\partial x} - \frac{\partial G}{\partial N_{Mix}}\frac{\partial N_{Mix}}{\partial x}$ equals $\alpha_1\left(W(H_F^{Skill})(1 - N_S(H_F^{Skill})\underline{t}_C) - W(\underline{H}_F)(1 - N_{Mix}\underline{t}_C)\right) - \left(((\alpha_1\alpha + \alpha_2)V'(N_{Mix})) - \alpha_1\delta W(\underline{H}_F)\underline{t}_C\right)\frac{W(\underline{H}_F)(1 - p_{Mix}a)\underline{t}_C}{\alpha V''(N_{Mix}(\underline{H}_F))}$. As long as the parents strictly prefer more grandchildren, then this expression is positive.

Proof of Lemma 2: It is never possible for both types of parents to both work and for some more able parents not to work, since the change in children's beliefs between working and not working that makes more able mothers indifferent between the two actions, will ensure that less able mothers strictly prefer not working. Likewise, if both types of parents work and some less able parents don't work, then children must infer in equilibrium that the non-working parents are less able, and this will ensure that all less able parents choose not to work. It is also impossible for both types of workers to work, and for no one to not work, as long as D1 holds, because the benefits from not working are strictly greater for the less able parents and as such, children will infer a non-working mother as having low workplace ability, which will ensure that less able parents will want to deviate. It is likewise impossible for all unable parents to work and all able parents to not work, because if the less able parents prefer working (given equilibrium beliefs) then the more able parents will prefer this as well. This rules out any equilibrium where less able parents work.

As long as $\theta(\delta W_F - H) > \Delta(1)$, , then more able mothers will prefer working and being known to be high ability to not working and being thought to be low ability. At these parameter values it is impossible to have a pooling equilibrium where more able parents don't work, since skilled $\Delta(p_0)$<$\Delta(1)$ and in such a pooling equilibrium, the benefit for skilled parents from deviating would be $\theta(\delta W_F - H) - \Delta(p_0) > 0$, which ensures deviation. It is also impossible to have a semi-pooling equilibrium where some skilled parents don't work, since the implied difference in parental welfare associated with the change in beliefs between working and not working, which we denote $\Delta(p_{pool})$ must lie between $\Delta(p_0)$ and $\Delta(1)$ and this cannot make the skilled parents indifferent between the two actions.

If $\theta(\delta W_F - H) < \Delta(p_0)$, then both types of parents will prefer pooling and not-working to working even if working is thought to indicate high level of ability, which it would have to given D1. At these parameter values, neither skilled nor less skilled parents would want to deviate by working because the welfare gain from switching beliefs relative to the pooling equilibrium are insufficiently small. At these parameter values, a pure separating equilibrium is impossible since $\Delta(p_0)<\Delta(1)$ which implies that the benefit for the skilled of imitating the less skilled is sufficiently large, so that deviation to not working would be optimal. Finally, a semi-pooling equilibrium is similarly impossible since the difference in beliefs between not working and working, denoted again $\Delta(p_{pool})$must again be greater than $\Delta(p_0)$ and that ensures that all of the skilled would prefer not working.

When $\Delta(1) > \theta(\delta W_F - H) > \Delta(p_0)$, pure pooling is impossible, since the skilled would like to deviate by working and forgoing $\Delta(p_0)$ to get $\theta(\delta W_F - H)$. Pure separating is similarly impossible since the more skilled would like to deviate from working to not working and gaining $\Delta(1$. The only remaining equilibrium is that some of the skilled imitate the non-skilled and choose to not work. In this case, $\Delta(p_{pool})$ must equal $\theta(\delta W_F - H)$.

In the pooling equilibrium, all children have the posterior that p_0 women are less able. In the separating equilibrium, a fraction p_0 of children believe that women are less able with probability $p_0 + \frac{\Delta^2}{3p_0}$ and a fraction $1 - p_0$ of children believe that women are less able with probability $p_0 - \frac{\Delta^2}{3(1-p_0)}$, and summing these quantities together yields that the average belief in the population is that on average, children believe that p_0 of women are less able in the workplace.

Finally, in the semi-pooling equilibrium, assume that a fraction γ of more able parents choose not to work, and if the posterior probability of having a less skilled parent conditional upon seeing that parent not work equals . $\frac{p_0}{p_0+\gamma(1-p_0)}$, then the belief that the women are less able is $p_0 - \frac{\Delta^2}{3(1-p_0)}$ for $(1-\gamma)(1-p_0)$ of the population and $p_0 + \frac{\Delta^2\left(\frac{p_0}{p_0+\gamma(1-p_0)} - p_0\right)}{3(1-p_0)p_0}$ for a fraction $p_0 + \gamma(1-p_0)$ of the population. Summing these together yields p_0.

Proof of Lemma 2: The first derivative of $\Delta(p, \hat{\alpha}_2)$ with respect to p equals $(\alpha_1\alpha + \hat{\alpha}_2)V'\left(N\left(p_0 + \frac{\Delta^2(p-p_0)}{3(1-p_0)p_0}\right)\right) - \alpha_1\delta W_F(1-p_P a)\underline{t}_C$ times the derivative of N with respect to p which is positive. The second derivative of this at its maximum is always negative, so it reaches a single maximum (although it is possible non-concave elsewhere). When $\hat{\alpha}_2 = 0$, the derivative must be negative since $\alpha V'\left(N\left(p_0+\frac{\Delta^2(p-p_0)}{3(1-p_0)p_0}\right)\right)=\left(1-\mu\left(p_0+\frac{\Delta^2(p-p_0)}{3(1-p_0)p_0}\right)a-(1-\mu)p_F a\right)\delta W_F \underline{t}_C<\alpha_1\delta W_F(1-p_P a)\underline{t}_C$. When $\hat{\alpha}_2=0$, $\Delta(p,\hat{\alpha}_2)$ is also negative since $(\alpha_1\alpha+\hat{\alpha}_2)V(N)-\alpha_1\delta W_F\left(1-\left(p_0-\frac{\Delta^2}{3(1-p_0)}\right)a\right)\underline{t}_C N$ is maximized as $N\left(p_0-\frac{\Delta^2}{3(1-p_0)}\right)$. When $\hat{\alpha}_2=\alpha_2$, by assumption, $\Delta(p,\hat{\alpha}_2)$ is strictly increasing with p and it is strictly positive. Using the notation that $p_F=p_0+\vartheta$, when

When $\hat{\alpha}_2=\alpha\alpha_1 a\frac{\frac{\Delta^2(\mu+(1-\mu)p_0)}{3p_0(1-p_0)}+(1-\mu)\vartheta}{1-ap_0-\mu a\frac{\Delta^2}{3p_0}-(1-\mu)a\vartheta}$, the maximum value of $\Delta(p,\hat{\alpha}_2)$ is reached at p=1, and for all lower values of $\hat{\alpha}_2$, the maximum value is below 1. But at that value $\Delta(p,\hat{\alpha}_2)$ equals $\left(\frac{1-ap_0+a\frac{\Delta^2}{3(1-p_0)}}{1-ap_0-\mu a\frac{\Delta^2}{3p_0}-(1-\mu)a\vartheta}\right)\alpha_1$ times $\alpha\left(V\left(N\left(p_0+\frac{\Delta^2}{3p_0}\right)\right)-V\left(N\left(p_0-\frac{\Delta^2}{3(1-p_0)}\right)\right)\right)$ minus $\delta W_F\left(1-ap_0-\mu a\frac{\Delta^2}{3p_0}-(1-\mu)a\vartheta\right)\underline{t}_C\left(N\left(p_0+\frac{\Delta^2}{3p_0}\right)-N\left(p_0-\frac{\Delta^2}{3(1-p_0)}\right)\right)$, and $\alpha V'\left(N\left(p_0-\frac{\Delta^2}{3(1-p_0)}\right)\right)=\delta W_F\left(1-ap_0-\mu a\frac{\Delta^2}{3p_0}-(1-\mu)a\vartheta\right)\underline{t}_C$. As V(.) is concave, $\alpha\left(V\left(N\left(p_0+\frac{\Delta^2}{3p_0}\right)\right)-V\left(N\left(p_0-\frac{\Delta^2}{3(1-p_0)}\right)\right)\right)$ is less than $\alpha V'\left(N\left(p_0-\frac{\Delta^2}{3(1-p_0)}\right)\right)$ times $N\left(p_0+\frac{\Delta^2}{3p_0}\right)-N\left(p_0-\frac{\Delta^2}{3(1-p_0)}\right)$ and thus the overall expression must be negative. This proves that whenever $\Delta(p,\hat{\alpha}_2)$ is positive, $\hat{\alpha}_2>\alpha\alpha_1 a\frac{\frac{\Delta^2(\mu+(1-\mu)p_0)}{3p_0(1-p_0)}+(1-\mu)\vartheta}{1-ap_0-\mu a\frac{\Delta^2}{3p_0}-(1-\mu)a\vartheta}$, and the slope of $\Delta(p,\hat{\alpha}_2)$ is positive with respect to p, at least up to the point where p=1.

Proof of Lemma 3: At $\hat{\alpha}_2=0$, then $\hat{\Delta}_1<0$, since $N\left(p_0-\frac{\Delta^2}{3(1-p_0)}\right)$, maximizes; $\alpha V(N)-\delta W_F\left(1-\left(p_0-\frac{\Delta^2}{3(1-p_0)}\right)a\right)\underline{t}_C N$, and thus $\alpha V\left(N\left(p_0+\frac{\Delta^2}{3p_0}\right)\right)-\delta W_F(1-p_{pA}a)\underline{t}_C N\left(p_0+\frac{\Delta^2}{3p_0}\right)<\alpha V\left(N\left(p_0-\frac{\Delta^2}{3(1-p_0)}\right)\right)-\delta W_F(1-p_{pA}a)\underline{t}_C N\left(p_0-\frac{\Delta^2}{3(1-p_0)}\right)$. As $\alpha V'(N)<\delta W_F\left(1-\left(p_0-\frac{\Delta^2}{3(1-p_0)}\right)a\right)\underline{t}_C$,, for all N> $N\left(p_0-\frac{\Delta^2}{3(1-p_0)}\right)$, it must be true that $\hat{\Delta}_1<\hat{\Delta}_2<0$. We also have assumed that when $\hat{\alpha}_2=\alpha_2$, $\hat{\Delta}_1>\hat{\Delta}_2>0$

The derivative of $\hat{\Delta}_1$ with respect to $\hat{\alpha}_2$ is positive and strictly greater than the derivative of $\hat{\Delta}_2$ with respect to $\hat{\alpha}_2$, which is also positive, which implies that $\Delta_1>\hat{\Delta}_1$ and $\Delta_2>\hat{\Delta}_2$. Also, there must exist a value of $\hat{\alpha}_2$, equal to

$$\frac{\alpha_1\left(\alpha\left(V\left(N\left(p_0-\frac{\Delta^2}{3(1-p_0)}\right)\right)-V\big(N(p_0)\big)\right)-\delta W_F(1-p_{pA}a)\underline{t}_C\left(N\left(p_0-\frac{\Delta^2}{3(1-p_0)}\right)-N(p_0)\right)\right)}{V\big(N(p_0)\big)-V\left(N\left(p_0-\frac{\Delta^2}{3(1-p_0)}\right)\right)}$$

denoted $\hat{\alpha}_{2,,=} < \alpha_2$ at which $\hat{\Delta}_1 = \hat{\Delta}_2$ and for all . $\hat{\alpha}_2 > \hat{\alpha}_{2,,=}$, $\hat{\Delta}_1 > \hat{\Delta}_2$, while for all $\hat{\alpha}_2 < \hat{\alpha}_{2,,=}$, $\hat{\Delta}_1 < \hat{\Delta}_2$. At $\hat{\alpha}_2 = \hat{\alpha}_{2,,=}$, $\hat{\Delta}_1 = \hat{\Delta}_2 > 0$ if and only if

$$\frac{N(p_0)-N\left(p_0-\frac{\Delta^2}{3(1-p_0)}\right)}{N\left(p_0+\frac{\Delta^2}{3p_0}\right)-N(p_0)} > \frac{V\big(N(p_0)\big)-V\left(N\left(p_0-\frac{\Delta^2}{3(1-p_0)}\right)\right)}{V\left(N\left(p_0+\frac{\Delta^2}{3p_0}\right)\right)-V\big(N(p_0)\big)}$$

But this can never hold since $V\big(N(p_0)\big)-V\left(N\left(p_0-\frac{\Delta^2}{3(1-p_0)}\right)\right) > V'\big(N(p_0)\big)\left(N(p_0)-N\left(p_0-\frac{\Delta^2}{3(1-p_0)}\right)\right)$ and $V\left(N\left(p_0+\frac{\Delta^2}{3p_0}\right)\right)-V\big(N(p_0)\big) < V'\big(N(p_0)\big)\left(N\left(p_0+\frac{\Delta^2}{3p_0}\right)-N(p_0)\right)$, so at $\hat{\alpha}_2 = \hat{\alpha}_{2,,=}$, $\hat{\Delta}_1 = \hat{\Delta}_2 < 0$.

Proof of Proposition 9: As above, it is impossible to have an equilibrium where both types of parents work and don't work, since the beliefs that would make one type of parents indifferent between the two action types will ensure that one type of parent will strictly prefer one of the two effort types. It is also impossible to have able parents working and less able parents mixing between working and not working, since the less able parents would always prefer not working and being known to be less able. It is also impossible to have a pooling equilibrium where both types of parents work, as long as D1 holds, since a deviation to not working will always be more likely to come from a less able parents, and therefore children should believe that this deviation comes from the less able and that is enough to ensure deviation by less able parents.

As such, there are only three possible types of equilibrium, for any value of $\hat{\alpha}_2$, separating where high ability parents work and low ability parents don't work, pure pooling where all parents don't work and semi-pooling where some high ability parents work and some don't work, and all low ability parents don't work.

If $(\delta W_F - H) > \Delta(1, \alpha_2)$, , and if children believe that able parents work and less able parents don't work, then it is optimal for parents to separate. For all values of $\hat{\alpha}_2$ between 0 and α_2, children believe that the optimal parental response is to separate. Hence a separating equilibrium with correct beliefs is an equilibrium. As above, this equilibrium will yield the correct beliefs on average, even if the children of more able parents overstate the ability of women in equilibrium and the children of less able parents understate the ability of women in equilibrium.

Pure pooling is not an equilibrium outcome, since $\theta(\delta W_F - H) > \Delta(1, \alpha_2)$, as high ability parents would always deviate. Moreover, in a semi-pooling equilibrium, the belief advantages to parents not-working must be less than $\Delta(1, \alpha_2)$, and this ensures skilled parents will want to work.

If $\theta(\delta W_F - H) < \Delta(1, \alpha_2)$, , then a pure separating equilibrium is not a possibility, since skilled parents will always deviate and try to pool with the less skilled. We first consider cases where $\theta(\delta W_F - H) > \Delta(p_0, \alpha_2)$, and classic pooling equilibrium is also impossible since skilled parents will want to deviate and work as long as they are thought to be a mixture of skilled and unskilled if they don't' work.

For values of $\hat{\alpha}_2$, close to but below α_2, children will continue to believe that it is optimal for parents to mix and they will have beliefs that will satisfy:

$$\begin{aligned}\theta(\delta W_F - H) = \alpha_1 \delta W_F \left(1 - \left(p_0 - \frac{\Delta^2}{3p_0}\right) a\right) \underline{t}_C \Bigg(N\left(p_0 - \frac{\Delta^2}{3(1-p_0)}\right) \\ - N\left(p_0 + \frac{\Delta^2(p_{Mix} - p_0)}{3(1-p_0)p_0}\right)\Bigg) \\ + (\alpha_1 \alpha + \hat{\alpha}_2)\left(V\left(N\left(p_0 + \frac{\Delta^2(p_{Mix} - p_0)}{3(1-p_0)p_0}\right)\right) - V\left(N\left(p_0 - \frac{\Delta^2}{3(1-p_0)}\right)\right)\right)\end{aligned}$$

While semi-pooling is optimal for parents with $\alpha_2 = \hat{\alpha}_2$, since $\hat{\alpha}_2 < \alpha_2$, these beliefs imply that

$$\begin{aligned}\theta(\delta W_F - H) < \alpha_1 \delta W_F \left(1 - \left(p_0 - \frac{\Delta^2}{3p_0}\right) a\right) \underline{t}_C \Bigg(N\left(p_0 - \frac{\Delta^2}{3(1-p_0)}\right) \\ - N\left(p_0 + \frac{\Delta^2(p_{Mix} - p_0)}{3(1-p_0)p_0}\right)\Bigg) \\ + (\alpha_1 \alpha + \alpha_2)\left(V\left(N\left(p_0 + \frac{\Delta^2(p_{Mix} - p_0)}{3(1-p_0)p_0}\right)\right) - V\left(N\left(p_0 - \frac{\Delta^2}{3(1-p_0)}\right)\right)\right)\end{aligned}$$

As such all parents don't work.

If $(\delta W_F - H) < \Delta(p_0, \alpha_2)$, , then parents will pool, and as long as $\theta(\delta W_F - H) < \Delta(p_0, \hat{\alpha}_2)$,), then children also believe that it is optimal for parents to pool. For lower value of $\hat{\alpha}_2$, children believe that skilled parents will not stick at that equilibrium, and as such they will believe that some skilled parents will work. But as they believe that the share of skilled parents not working is actually less than it actually is, that only strengthens the incentive not work for skilled parents and reinforces the pooling equilibrium.

Proof of Proposition 10: When $\Delta(1,\alpha_2) > \theta(\delta W_F - H) > \Delta(p_0,\alpha_2)$, and $\alpha_2 > \hat{\alpha}_2$, , children's beliefs must satisfy

$$\theta(\delta W_F - H) = (\alpha_1\alpha + \hat{\alpha}_2)\left(V\left(N\left(p_0 + \frac{\Delta^2(p-p_0)}{3(1-p_0)p_0}\right)\right) - V\left(N\left(p_0 - \frac{\Delta^2}{3(1-p_0)}\right)\right)\right) - \alpha_1\delta W_F(1-p_P a)\underline{t}_C\left(N\left(p_0 + \frac{\Delta^2(p-p_0)}{3(1-p_0)p_0}\right) - N\left(p_0 - \frac{\Delta^2}{3(1-p_0)}\right)\right)$$

The left hand side of the equation is strictly increasing with p, as proven in Lemma 2, which implies that the value of p rises as $\hat{\alpha}_2$ falls, and eventually converges to 1 as $\hat{\alpha}_2$ goes to

$$\frac{\theta(\delta W_F - H) + \alpha_1\delta W_F(1-p_P a)\underline{t}_C\left(N\left(p_0 + \frac{\Delta^2}{3p_0}\right) - N\left(p_0 - \frac{\Delta^2}{3(1-p_0)}\right)\right)}{V\left(N\left(p_0 + \frac{\Delta^2}{3p_0}\right)\right) - V\left(N\left(p_0 - \frac{\Delta^2}{3(1-p_0)}\right)\right)} - \alpha_1\alpha$$

which equals $\hat{\alpha}_2^*$. At that point, the value of $\Delta(1,\hat{\alpha}_2) = \theta(\delta W_F - H)$, and for all values of p below1, $\theta(\delta W_F - H) > \Delta(p,\hat{\alpha}_2)$. For all lower levels of $\hat{\alpha}_2$, children believe that all skilled parents would work and therefore all children believe that their own parents are unskilled which leads to a universal belief women are less able with probability $p_0 + \frac{\Delta^2}{3p_0}$.

When $\theta(\delta W_F - H) < \Delta(p_0,\alpha_2)$, then for high enough values of $\hat{\alpha}_2$, children correctly perceive that they are in a pure pooling equilibrium and hold correct beliefs. For values of $\hat{\alpha}_2$ below the point where $\theta(\delta W_F - H) < \Delta(p_0,\hat{\alpha}_2)$, children believe that they are a semi-pooling equilibrium. That point occurs when $\hat{\alpha}_2$ equals

$\frac{\theta(\delta W_F - H) + \alpha_1\delta W_F(1-p_P a)\underline{t}_C\left(N(p_0) - N\left(p_0 - \frac{\Delta^2}{3(1-p_0)}\right)\right)}{V(N(p_0)) - V\left(N\left(p_0 - \frac{\Delta^2}{3(1-p_0)}\right)\right)} - \alpha_1\alpha$ or $\hat{\alpha}_2^{**}$. For lower levels of $\hat{\alpha}_2$, children believe that there are in semi-pooling equilibrium as long as $\hat{\alpha}_2$ is greater than

$$\frac{\theta(\delta W_F - H) + \alpha_1\delta W_F(1-p_P a)\underline{t}_C\left(N\left(p_0 + \frac{\Delta^2}{3p_0}\right) - N\left(p_0 - \frac{\Delta^2}{3(1-p_0)}\right)\right)}{V\left(N\left(p_0 + \frac{\Delta^2}{3p_0}\right)\right) - V\left(N\left(p_0 - \frac{\Delta^2}{3(1-p_0)}\right)\right)} - \alpha_1\alpha$$

For lower values of $\hat{\alpha}_2$, children believe that they are in a pure separating equilibrium where there is a universal belief women are less able with probability $p_0 + \frac{\Delta^2}{3p_0}$.

Since the value of $\hat{\alpha}_2^*$ is defined so that $\theta(\delta W_F - H) = \Delta(1, \hat{\alpha}_2^*)$ and the value of $\hat{\alpha}_2^{**}$ is defined so that $\theta(\delta W_F - H) = \Delta(p_0, \hat{\alpha}_2^*)$ it follows that both $\hat{\alpha}_2^*$ and $\hat{\alpha}_2^{**}$ are increasing with θ and decreasing with H.

$$\begin{aligned}\theta(\delta W_F - H) &= \Delta(p, \hat{\alpha}_2^*) \\ &= (\alpha_1 \alpha + \hat{\alpha}_2^*)\left(V\left(N\left(p_0 + \frac{\Delta^2}{3p_0}\right)\right) - V\left(N\left(p_0 - \frac{\Delta^2}{3(1-p_0)}\right)\right)\right) \\ &\quad - \alpha_1 \delta W_F (1 - p_P a)\underline{t}_C\left(N\left(p_0 + \frac{\Delta^2}{3(1-p_0)p_0}\right) - N\left(p_0 - \frac{\Delta^2}{3(1-p_0)}\right)\right)\end{aligned}$$

It follows that $\hat{\alpha}_2^*$ is increasing with θ and decreasing with H. In any semi-pooling equilibrium

$$\begin{aligned}\theta(\delta W_F - H) = (\alpha_1 \alpha + \hat{\alpha}_2)&\left(V\left(N\left(p_0 + \frac{\Delta^2(p-p_0)}{3(1-p_0)p_0}\right)\right) - V\left(N\left(p_0 - \frac{\Delta^2}{3(1-p_0)}\right)\right)\right) \\ &- \alpha_1 \delta W_F (1 - p_P a)\underline{t}_C\left(N\left(p_0 + \frac{\Delta^2(p-p_0)}{3(1-p_0)p_0}\right) - N\left(p_0 - \frac{\Delta^2}{3(1-p_0)}\right)\right)\end{aligned}$$

This implies that p must also be increasing with θ and decreasing with H.

Proof of Proposition 11: If $t_0 < \psi$, then lifetime utility equals: $\delta W_0\big((1-T)(1-ap_0) - a(1-p_0)(\psi + t_0)\big) + \alpha V(g(t_0)T/\underline{t}_C)$, which is clearly decreasing in t_0. The first order condition is that $\delta W_0(1-p_0 a) = \alpha V'(T/\underline{t}_C)/\underline{t}_C$, and we let T_I denote the level of fertility that occurs when the person is fundamentally ignorant about their ability level. Lifetime utility equals: $\delta W_0\big((1-T_I)(1-ap_0) - a(1-p_0)\psi\big) + \alpha V(T_I/\underline{t}_C)$.

If $V\left(\frac{g(\psi)T_L}{\underline{t}_C}\right) = \left(\frac{g(\psi)T_L}{\underline{t}_C}\right)^{\mu}$, then $T_I = \underline{t}_C^{\frac{-\mu}{1-\mu}}\left(\frac{\mu\alpha}{\delta W_0(1-p_0 a)}\right)^{\frac{1}{1-\mu}}$ and lifetime utility equals

$$\delta W_0(1 - ap_0 - a(1-p_0)\psi) + \alpha^{\frac{1}{1-\mu}}\underline{t}_C^{\frac{-\mu}{1-\mu}}\big(\delta W_0(1-p_0 a)\big)^{\frac{-\mu}{1-\mu}}\left(\mu^{\frac{\mu}{1-\mu}} - \mu^{\frac{1}{1-\mu}}\right).$$

When $t_0 \geq \psi$, then the individual learns their ability level before choosing the number of children. Welfare if the person knows themselves to be high ability will equal $\delta W_0\big((1-T) - a(\psi + t_0)\big) + \alpha V(g(t_0)T/\underline{t}_C)$, which is clearly decreasing in t_0, so the individual will set $t_0 = \psi$ and choose T so that $\delta W_0 = \alpha g(\psi) V'(g(\psi)T/\underline{t}_C)/\underline{t}_C$. If the individual learns that they are low ability their lifetime welfare is $\delta W_0\big((1-T)(1-a)\big) + \alpha V(g(t_0)T/\underline{t}_C)$, which is also decreasing in t_0, so that the individual sets $t_0 = \psi$, and sets T so that $\delta W_0(1-a) = \alpha g(\psi)V'(g(\psi)T/\underline{t}_C)/\underline{t}_C$. We let T_H denote the higher level of fertility that occurs when the person learns themselves to be less able and T_L denote the lower level of fertility that occurs when the person learns themselves to be more able.

If $V\left(\frac{g(\psi)T_L}{\underline{t}_C}\right) = \left(\frac{g(\psi)T_L}{\underline{t}_C}\right)^{\mu}$, then $T_H = g(\psi)^{\frac{\mu}{1-\mu}}\underline{t}_C^{\frac{-\mu}{1-\mu}}\left(\frac{\mu\alpha}{\delta W_0(1-a)}\right)^{\frac{1}{1-\mu}}$ and $T_L = g(\psi)^{\frac{\mu}{1-\mu}}\underline{t}_C^{\frac{-\mu}{1-\mu}}\left(\frac{\mu\alpha}{\delta W_0}\right)^{\frac{1}{1-\mu}}$. Lifetime utility for the more able individuals equals $\delta W_0(1-2a\psi) + g(\psi)^{\frac{\mu}{1-\mu}}\alpha^{\frac{1}{1-\mu}}\underline{t}_C^{\frac{-\mu}{1-\mu}}(\delta W_0)^{\frac{-\mu}{1-\mu}}\left(\mu^{\frac{\mu}{1-\mu}} - \mu^{\frac{1}{1-\mu}}\right)$. Lifetime utility for the less able individuals equal $\delta W_0(1-a) + g(\psi)^{\frac{\mu}{1-\mu}}\alpha^{\frac{1}{1-\mu}}\underline{t}_C^{\frac{-\mu}{1-\mu}}\left(\delta W_0(1-a)\right)^{\frac{-\mu}{1-\mu}}\left(\mu^{\frac{\mu}{1-\mu}} - \mu^{\frac{1}{1-\mu}}\right)$. Total expected utility equals

$\delta W_0(1-p_0a-2(1-p_0)a\psi) + \left(1-p_0+p_0(1-a)^{\frac{-\mu}{1-\mu}}\right)g(\psi)^{\frac{\mu}{1-\mu}}\alpha^{\frac{1}{1-\mu}}\underline{t}_C^{\frac{-\mu}{1-\mu}}(\delta W_0)^{\frac{-\mu}{1-\mu}}\left(\mu^{\frac{\mu}{1-\mu}} - \mu^{\frac{1}{1-\mu}}\right)$.

$\delta W_0(1-ap_0-a(1-p_0)\psi) + \alpha^{\frac{1}{1-\mu}}\underline{t}_C^{\frac{-\mu}{1-\mu}}\left(\delta W_0(1-p_0a)\right)^{\frac{-\mu}{1-\mu}}\left(\mu^{\frac{\mu}{1-\mu}} - \mu^{\frac{1}{1-\mu}}\right)$.

Lifetime utility in the case where the individual sets $t_0 = 0$, will be higher if and only if

$\psi\left(\frac{t_C}{\alpha}\right)^{\frac{\mu}{1-\mu}}(\delta W_0)^{\frac{1}{1-\mu}} > \frac{1}{1-p_0}\left(\left(1-p_0+p_0(1-a)^{\frac{-\mu}{1-\mu}}\right)g(\psi)^{\frac{\mu}{1-\mu}} - (1-p_0a)^{\frac{-\mu}{1-\mu}}\right)\left(\mu^{\frac{\mu}{1-\mu}} - \mu^{\frac{1}{1-\mu}}\right)$. The left hand side of the inequality is obviously negative if $g(\psi)$ equals zero. If $g(\psi)$ equals one, then Jensen's inequality ensures that $1-p_0+p_0(1-a)^{\frac{-\mu}{1-\mu}} - (1-p_0a)^{\frac{-\mu}{1-\mu}}$ is positive (since $(1-x)^{\frac{-\mu}{1-\mu}}$ is convex).

Hence as long as $\left(\frac{t_C}{\alpha}\right)^{\frac{\mu}{1-\mu}}(\delta W_0)^{\frac{1}{1-\mu}}$ is sufficiently small, the inequality must hold. To get concrete bounds on this quantity, assumem that $\frac{\mu}{1-\mu} = 1$, then the condition becomes:
$4\psi\left(\frac{t_C}{\alpha}\right)(\delta W_0)^2 < 1 - \frac{ap_0}{(1-a)(1-ap_0)}$.

When this condition holds, there exists a value of $g(\psi)$ for which women are indifferent between waiting and learning and for all higher values of $g(\psi)$ they strictly prefer waiting and for all lower values of $g(\psi)$ they strictly prefer having children immediately.

Define $g(\psi)^*$ as the value of $g(\psi)$ at which:

$\psi\left(\frac{\underline{t}_C}{\alpha}\right)^{\frac{\mu}{1-\mu}}(\delta W_0)^{\frac{1}{1-\mu}} = \frac{1}{1-p_0}\left(\left(1-p_0+p_0(1-a)^{\frac{-\mu}{1-\mu}}\right)g(\psi)^{*\frac{\mu}{1-\mu}} - (1-p_0a)^{\frac{-\mu}{1-\mu}}\right)\left(\mu^{\frac{\mu}{1-\mu}} - \mu^{\frac{1}{1-\mu}}\right)$. The right hand side of the equation is increasing with $g(\psi)^*$, which means that $g(\psi)^*$ is rising with $\underline{t}_C$, falling with α and rising with δ and W_0.

The derivative of $\left(\left(1-p_0+p_0(1-a)^{\frac{-\mu}{1-\mu}}\right)g(\psi)^{*\frac{\mu}{1-\mu}} - (1-p_0a)^{\frac{-\mu}{1-\mu}}\right)$ with respect to a is$\frac{\mu}{1-\mu}\left(p_0(1-a)^{\frac{-1}{1-\mu}}g(\psi)^{*\frac{\mu}{1-\mu}} - (1-p_0a)^{\frac{-1}{1-\mu}}\right)$, which is negative if and only if $p_0^{1-\mu}(1-p_0a)g(\psi)^{*\mu} < (1-a)$. When p_0 is close to zero or one, this must hold.

Finally we consider the impact of p_0. Treating everything else as a constant, the equation can be rewritten: $K = \left(\left(1+\frac{p_0}{1-p_0}(1-a)^{\frac{-\mu}{1-\mu}}\right)g(\psi)^{*\frac{\mu}{1-\mu}} - \frac{(1-p_0a)^{\frac{-\mu}{1-\mu}}}{1-p_0}\right)$.. The derivative of $g(\psi)^*$ with respect to $p_0$is positive if and only if the derivative of the right hand side of that equation with respect to p_0, which equals $\frac{(1-a)^{\frac{-\mu}{1-\mu}}g(\psi)^{*\frac{\mu}{1-\mu}} - \frac{\mu}{1-\mu}(1-p_0a)^{\frac{-1}{1-\mu}}a(1-p_0)+(1-p_0a)^{-\frac{\mu}{1-\mu}}}{(1-p_0)^2}$, is negative. For this to be positive it must be that$(1-a)^{\frac{-\mu}{1-\mu}}g(\psi)^{*\frac{\mu}{1-\mu}}(1-p_0a)^{\frac{1}{1-\mu}} + (1-p_0a) > \frac{\mu}{1-\mu}a(1-p_0)$. For p_0 near one, this clearly holds because the right hand side of the inequality converges to zero. When $\mu \leq .5$, the inequality always holds.

Proof of Proposition 12: For decision-makers who wait, persuasion will have no effect, since the decision-maker will know the truth before they make child-related investments. When the decision-maker has children before work, then the number of children equals $\underline{t}_C^{\frac{-\mu}{1-\mu}}\left(\frac{\mu\alpha}{\delta W_0(1-p_0a)}\right)^{\frac{1}{1-\mu}}$, which is clearly increasing in p_0.

When the decision-maker is on the margin, then an increase in p_0 can induce them to have more children only if it raises $g(\psi)^*$, which is far from guaranteed. Moreover, the decision-maker in expectation will have more children if she doesn't wait if and only if $(1-p_0a)^{\frac{-1}{1-\mu}} >$

$p(1-a)^{\frac{-1}{1-\mu}} + (1-p)$, where p refers to the real probability of being less able. If $p_0 = p$, then this condition can never hold since $(1-x)^{\frac{-1}{1-\mu}}$ is convex in p. If however, $p_0 > p$, then this condition may hold.

Figure 1: Opinions about Women Working by Year of Birth (Multiple Years)

Figure 2: Women Working and Children's Welfare (Multiple Years)

Figure 3: Mean are Better at Politics (Multiple Years)

Figure 4: Are Women Good for the Military (1983)

Figure 5: Men Earn More Because They Work Harder (1996)

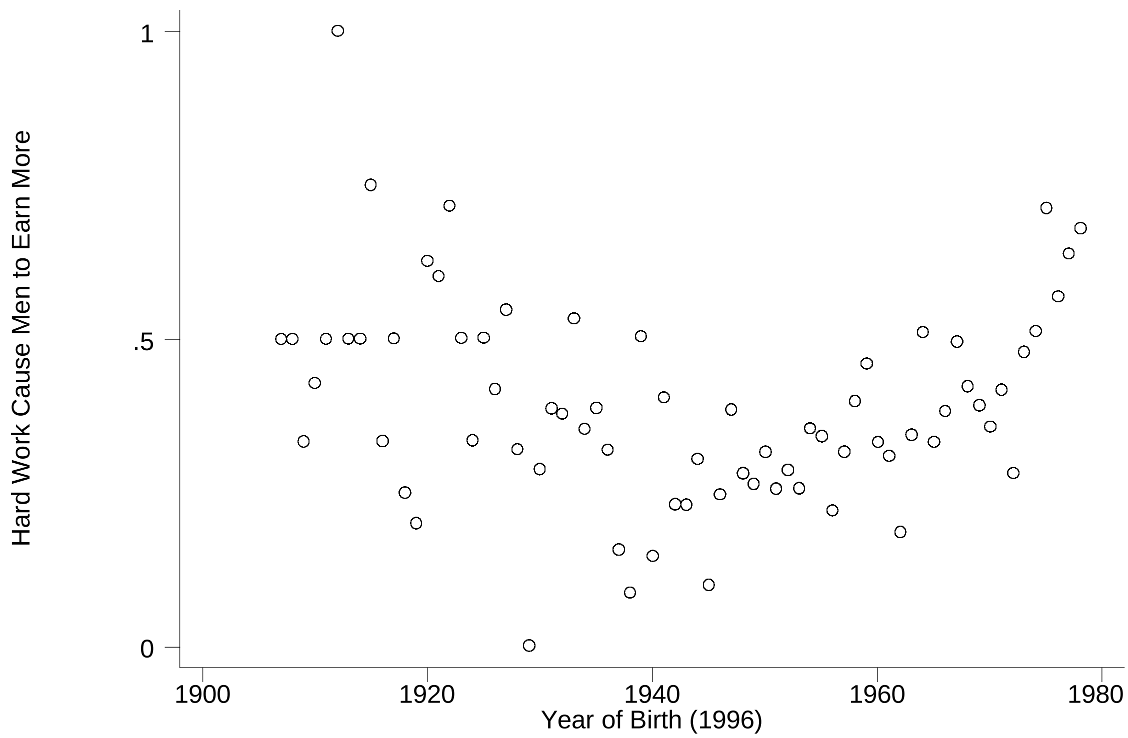

Figure 6: Gender Aids at Work (1987)

Figure 7: Jobs are Worse for Women than Men (1985)