This is a famous (or infamous) talk in which the president of Harvard argued that biological differences may be a key contributing factor to women's relative lack of success in the sciences and math.

Diversifying the Science & Engineering Workforce
Lawrence H. Summers Cambridge, Mass. January 14, 2005

I asked Richard, when he invited me to come here and speak, whether he wanted an institutional talk about Harvard's policies toward diversity or whether he wanted some questions asked and some attempts at provocation, because I was willing to do the second and didn't feel like doing the first. And so we have agreed that I am speaking unofficially and not using this as an occasion to lay out the many things we're doing at Harvard to promote the crucial objective of diversity. There are many aspects of the problems you're discussing and it seems to me they're all very important from a national point of view. I'm going to confine myself to addressing one portion of the problem, or of the challenge we're discussing, which is the issue of women's representation in tenured positions in science and engineering at top universities and research institutions, not because that's necessarily the most important problem or the most interesting problem, but because it's the only one of these problems that I've made an effort to think in a very serious way about. The other prefatory comment that I would make is that I am going to, until most of the way through, attempt to adopt an entirely positive, rather than normative approach, and just try to think about and offer some hypotheses as to why we observe what we observe without seeing this through the kind of judgmental tendency that inevitably is connected with all our common goals of equality. It is after all not the case that the role of women in science is the only example of a group that is significantly underrepresented in an important activity and whose underrepresentation contributes to a shortage of role models for others who are considering being in that group. To take a set of diverse examples, the data will, I am confident, reveal that Catholics are substantially underrepresented in investment banking, which is an enormously high-paying profession in our society; that white men are very substantially underrepresented in the National Basketball Association; and that Jews are very substantially underrepresented in farming and in agriculture. These are all phenomena in which one observes underrepresentation, and I think it's important to try to think systematically and clinically about the reasons for underrepresentation.

There are three broad hypotheses about the sources of the very substantial disparities that this conference's papers document and have been documented before with respect to the presence of women in high-end scientific professions. One is what I would call the-I'll explain each of these in a few moments and comment on how important I think they are-the first is what I call the high-powered job hypothesis. The second is what I would call different availability of aptitude at the high end, and the third is what I would call different socialization and patterns of discrimination in a search. And in my own view, their importance probably ranks in exactly the order that I just described.

Maybe it would be helpful to just, for a moment, broaden the problem, or the issue, beyond science and engineering. I've had the opportunity to discuss questions like this with chief executive officers at major corporations, the managing partners of large law firms, the directors of prominent teaching hospitals, and with the leaders of other prominent professional service organizations, as well as with colleagues in higher education. In all of those groups, the story is fundamentally the same. Twenty or twenty-five years ago, we started to see very substantial increases in the number of women who were in graduate school in this field. Now the people who went to graduate school when that started are forty, forty-five, fifty years old. If you look at the top cohort in our activity, it is not only nothing like fifty-fifty, it is nothing like what we thought it was when we started having a third of the

women, a third of the law school class being female, twenty or twenty-five years ago. And the relatively few women who are in the highest ranking places are disproportionately either unmarried or without children, with the emphasis differing depending on just who you talk to. And that is a reality that is present and that one has exactly the same conversation in almost any high-powered profession. What does one make of that? I think it is hard-and again, I am speaking completely descriptively and non-normatively-to say that there are many professions and many activities, and the most prestigious activities in our society expect of people who are going to rise to leadership positions in their forties near total commitments to their work. They expect a large number of hours in the office, they expect a flexibility of schedules to respond to contingency, they expect a continuity of effort through the life cycle, and they expect-and this is harder to measure-but they expect that the mind is always working on the problems that are in the job, even when the job is not taking place. And it is a fact about our society that that is a level of commitment that a much higher fraction of married men have been historically prepared to make than of married women. That's not a judgment about how it should be, not a judgment about what they should expect. But it seems to me that it is very hard to look at the data and escape the conclusion that that expectation is meeting with the choices that people make and is contributing substantially to the outcomes that we observe. One can put it differently. Of a class, and the work that Claudia Goldin and Larry Katz are doing will, I'm sure, over time, contribute greatly to our understanding of these issues and for all I know may prove my conjectures completely wrong. Another way to put the point is to say, what fraction of young women in their mid-twenties make a decision that they don't want to have a job that they think about eighty hours a week. What fraction of young men make a decision that they're unwilling to have a job that they think about eighty hours a week, and to observe what the difference is. And that has got to be a large part of what is observed. Now that begs entirely the normative questions-which I'll get to a little later-of, is our society right to expect that level of effort from people who hold the most prominent jobs? Is our society right to have familial arrangements in which women are asked to make that choice and asked more to make that choice than men? Is our society right to ask of anybody to have a prominent job at this level of intensity, and I think those are all questions that I want to come back to. But it seems to me that it is impossible to look at this pattern and look at its pervasiveness and not conclude that something of the sort that I am describing has to be of significant importance. To buttress conviction and theory with anecdote, a young woman who worked very closely with me at the Treasury and who has subsequently gone on to work at Google highly successfully, is a 1994 graduate of Harvard Business School. She reports that of her first year section, there were twenty-two women, of whom three are working full time at this point. That may, the dean of the Business School reports to me, that that is not an implausible observation given their experience with their alumnae. So I think in terms of positive understanding, the first very important reality is just what I would call the, who wants to do high-powered intense work?

The second thing that I think one has to recognize is present is what I would call the combination of, and here, I'm focusing on something that would seek to answer the question of why is the pattern different in science and engineering, and why is the representation even lower and more problematic in science and engineering than it is in other fields. And here, you can get a fair distance, it seems to me, looking at a relatively simple hypothesis. It does appear that on many, many different human attributes-height, weight, propensity for criminality, overall IQ, mathematical ability, scientific ability-there is relatively clear evidence that whatever the difference in means-which can be debated-there is a difference in the standard deviation, and variability of a male and a female population. And that is true

with respect to attributes that are and are not plausibly, culturally determined. If one supposes, as I think is reasonable, that if one is talking about physicists at a top twenty-five research university, one is not talking about people who are two standard deviations above the mean. And perhaps it's not even talking about somebody who is three standard deviations above the mean. But it's talking about people who are three and a half, four standard deviations above the mean in the one in 5,000, one in 10,000 class. Even small differences in the standard deviation will translate into very large differences in the available pool substantially out. I did a very crude calculation, which I'm sure was wrong and certainly was unsubtle, twenty different ways. I looked at the Xie and Shauman paper-looked at the book, rather-looked at the evidence on the sex ratios in the top 5% of twelfth graders. If you look at those-they're all over the map, depends on which test, whether it's math, or science, and so forth-but 50% women, one woman for every two men, would be a high-end estimate from their estimates. From that, you can back out a difference in the implied standard deviations that works out to be about 20%. And from that, you can work out the difference out several standard deviations. If you do that calculation-and I have no reason to think that it couldn't be refined in a hundred ways-you get five to one, at the high end. Now, it's pointed out by one of the papers at this conference that these tests are not a very good measure and are not highly predictive with respect to people's ability to do that. And that's absolutely right. But I don't think that resolves the issue at all. Because if my reading of the data is right-it's something people can argue about-that there are some systematic differences in variability in different populations, then whatever the set of attributes are that are precisely defined to correlate with being an aeronautical engineer at MIT or being a chemist at Berkeley, those are probably different in their standard deviations as well. So my sense is that the unfortunate truth-I would far prefer to believe something else, because it would be easier to address what is surely a serious social problem if something else were true-is that the combination of the high-powered job hypothesis and the differing variances probably explains a fair amount of this problem.

There may also be elements, by the way, of differing, there is some, particularly in some attributes, that bear on engineering, there is reasonably strong evidence of taste differences between little girls and little boys that are not easy to attribute to socialization. I just returned from Israel, where we had the opportunity to visit a kibbutz, and to spend some time talking about the history of the kibbutz movement, and it is really very striking to hear how the movement started with an absolute commitment, of a kind one doesn't encounter in other places, that everybody was going to do the same jobs. Sometimes the women were going to fix the tractors, and the men were going to work in the nurseries, sometimes the men were going to fix the tractors and the women were going to work in the nurseries, and just under the pressure of what everyone wanted, in a hundred different kibbutzes, each one of which evolved, it all moved in the same direction. So, I think, while I would prefer to believe otherwise, I guess my experience with my two and a half year old twin daughters who were not given dolls and who were given trucks, and found themselves saying to each other, look, daddy truck is carrying the baby truck, tells me something. And I think it's just something that you probably have to recognize. There are two other hypotheses that are all over. One is socialization. Somehow little girls are all socialized towards nursing and little boys are socialized towards building bridges. No doubt there is some truth in that. I would be hesitant about assigning too much weight to that hypothesis for two reasons. First, most of what we've learned from empirical psychology in the last fifteen years has been that people naturally attribute things to socialization that are in fact not attributable to socialization. We've been astounded by the results of separated twins studies. The confident assertions that autism was a reflection of parental characteristics that

were absolutely supported and that people knew from years of observational evidence have now been proven to be wrong. And so, the human mind has a tendency to grab to the socialization hypothesis when you can see it, and it often turns out not to be true. The second empirical problem is that girls are persisting longer and longer. When there were no girls majoring in chemistry, when there were no girls majoring in biology, it was much easier to blame parental socialization. Then, as we are increasingly finding today, the problem is what's happening when people are twenty, or when people are twenty-five, in terms of their patterns, with which they drop out. Again, to the extent it can be addressed, it's a terrific thing to address.

The most controversial in a way, question, and the most difficult question to judge, is what is the role of discrimination? To what extent is there overt discrimination? Surely there is some. Much more tellingly, to what extent are there pervasive patterns of passive discrimination and stereotyping in which people like to choose people like themselves, and the people in the previous group are disproportionately white male, and so they choose people who are like themselves, who are disproportionately white male. No one who's been in a university department or who has been involved in personnel processes can deny that this kind of taste does go on, and it is something that happens, and it is something that absolutely, vigorously needs to be combated. On the other hand, I think before regarding it as pervasive, and as the dominant explanation of the patterns we observe, there are two points that should make one hesitate. The first is the fallacy of composition. No doubt it is true that if any one institution makes a major effort to focus on reducing stereotyping, on achieving diversity, on hiring more people, no doubt it can succeed in hiring more. But each person it hires will come from a different institution, and so everyone observes that when an institution works very hard at this, to some extent they are able to produce better results. If I stand up at a football game and everybody else is sitting down, I can see much better, but if everybody stands up, the views may get a little better, but they don't get a lot better. And there's a real question as to how plausible it is to believe that there is anything like half as many people who are qualified to be scientists at top ten schools and who are now not at top ten schools, and that's the argument that one has to make in thinking about this as a national problem rather than an individual institutional problem. The second problem is the one that Gary Becker very powerfully pointed out in addressing racial discrimination many years ago. If it was really the case that everybody was discriminating, there would be very substantial opportunities for a limited number of people who were not prepared to discriminate to assemble remarkable departments of high quality people at relatively limited cost simply by the act of their not discriminating, because of what it would mean for the pool that was available. And there are certainly examples of institutions that have focused on increasing their diversity to their substantial benefit, but if there was really a pervasive pattern of discrimination that was leaving an extraordinary number of high-quality potential candidates behind, one suspects that in the highly competitive academic marketplace, there would be more examples of institutions that succeeded substantially by working to fill the gap. And I think one sees relatively little evidence of that. So my best guess, to provoke you, of what's behind all of this is that the largest phenomenon, by far, is the general clash between people's legitimate family desires and employers' current desire for high power and high intensity, that in the special case of science and engineering, there are issues of intrinsic aptitude, and particularly of the variability of aptitude, and that those considerations are reinforced by what are in fact lesser factors involving socialization and continuing discrimination. I would like nothing better than to be proved wrong, because I would like nothing better than for these problems to be addressable simply by everybody understanding what they are, and working very hard to address them.

What's to be done? And what further questions should one know the answers to? Let me take a second, first to just remark on a few questions that it seems to me are ripe for research, and for all I know, some of them have been researched. First, it would be very useful to know, with hard data, what the quality of marginal hires are when major diversity efforts are mounted. When major diversity efforts are mounted, and consciousness is raised, and special efforts are made, and you look five years later at the quality of the people who have been hired during that period, how many are there who have turned out to be much better than the institutional norm who wouldn't have been found without a greater search. And how many of them are plausible compromises that aren't unreasonable, and how many of them are what the right-wing critics of all of this suppose represent clear abandonments of quality standards. I don't know the answer, but I think if people want to move the world on this question, they have to be willing to ask the question in ways that could face any possible answer that came out. Second, and by the way, I think a more systematic effort to look at citation records of male and female scholars in disciplines where citations are relatively well-correlated with academic rank and with people's judgments of quality would be very valuable. Of course, most of the critiques of citations go to reasons why they should not be useful in judging an individual scholar. Most of them are not reasons why they would not be useful in comparing two large groups of scholars and so there is significant potential, it seems to me, for citation analysis in this regard. Second, what about objective versus subjective factors in hiring? I've been exposed, by those who want to see the university hiring practices changed to favor women more and to assure more diversity, to two very different views. One group has urged that we make the processes consistently more clear-cut and objective, based on papers, numbers of papers published, numbers of articles cited, objectivity, measurement of performance, no judgments of potential, no reference to other things, because if it's made more objective, the subjectivity that is associated with discrimination and which invariably works to the disadvantage of minority groups will not be present. I've also been exposed to exactly the opposite view, that those criteria and those objective criteria systematically bias the comparisons away from many attributes that those who contribute to the diversity have: a greater sense of collegiality, a greater sense of institutional responsibility. Somebody ought to be able to figure out the answer to the question of, if you did it more objectively versus less objectively, what would happen. Then you can debate whether you should or whether you shouldn't, if objective or subjective is better. But that question ought to be a question that has an answer, that people can find. Third, the third kind of question is, what do we know about search procedures in universities? Is it the case that more systematic comprehensive search processes lead to minority group members who otherwise would have not been noticed being noticed? Or does fetishizing the search procedure make it very difficult to pursue the targets of opportunity that are often available arising out of particular family situations or particular moments, and does fetishizing and formalizing search procedures further actually work to the disadvantage of minority group members. Again, everybody's got an opinion; I don't think anybody actually has a clue as to what the answer is. Fourth, what do we actually know about the incidence of financial incentives and other support for child care in terms of what happens to people's career patterns. I've been struck at Harvard that there's something unfortunate and ironic about the fact that if you're a faculty member and you have a kid who's 18 who goes to college, we in effect, through an interest-free loan, give you about $9,000. If you have a six-year-old, we give you nothing. And I don't think we're very different from most other universities in this regard, but there is something odd about that strategic choice, if the goal is to recruit people to come to the university. But I don't think we know much about the child care issue. The fifth question-which it seems to me would be

useful to study and to actually learn the answer to-is what do we know, or what can we learn, about the costs of career interruptions. There is something we would like to believe. We would like to believe that you can take a year off, or two years off, or three years off, or be half-time for five years, and it affects your productivity during the time, but that it really doesn't have any fundamental effect on the career path. And a whole set of conclusions would follow from that in terms of flexible work arrangements and so forth. And the question is, in what areas of academic life and in what ways is it actually true. Somebody reported to me on a study that they found, I don't remember who had told me about this-maybe it was you, Richard-that there was a very clear correlation between the average length of time, from the time a paper was cited. That is, in fields where the average papers cited had been written nine months ago, women had a much harder time than in fields where the average thing cited had been written ten years ago. And that is suggestive in this regard. On the discouraging side of it, someone remarked once that no economist who had gone to work at the President's Council of Economic Advisors for two years had done highly important academic work after they returned. Now, I'm sure there are counterexamples to that, and I'm sure people are kind of processing that Tobin's Q is the best-known counterexample to that proposition, and there are obviously different kinds of effects that happen from working in Washington for two years. But it would be useful to explore a variety of kinds of natural interruption experiments, to see what actual difference it makes, and to see whether it's actually true, and to see in what ways interruptions can be managed, and in what fields it makes a difference. I think it's an area in which there's conviction but where it doesn't seem to me there's an enormous amount of evidence. What should we all do? I think the case is overwhelming for employers trying to be the [unintelligible] employer who responds to everybody else's discrimination by competing effectively to locate people who others are discriminating against, or to provide different compensation packages that will attract the people who would otherwise have enormous difficulty with child care. I think a lot of discussion of issues around child care, issues around extending tenure clocks, issues around providing family benefits, are enormously important. I think there's a strong case for monitoring and making sure that searches are done very carefully and that there are enough people looking and watching that that pattern of choosing people like yourself is not allowed to take insidious effect. But I think it's something that has to be done with very great care because it slides easily into pressure to achieve given fractions in given years, which runs the enormous risk of people who were hired because they were terrific being made to feel, or even if not made to feel, being seen by others as having been hired for some other reason. And I think that's something we all need to be enormously careful of as we approach these issues, and it's something we need to do, but I think it's something that we need to do with great care.

Let me just conclude by saying that I've given you my best guesses after a fair amount of reading the literature and a lot of talking to people. They may be all wrong. I will have served my purpose if I have provoked thought on this question and provoked the marshaling of evidence to contradict what I have said. But I think we all need to be thinking very hard about how to do better on these issues and that they are too important to sentimentalize rather than to think about in as rigorous and careful ways as we can. That's why I think conferences like this are very, very valuable. Thank you.