NOTES FOR A THEORY OF EXCELLENCE

Elisabetta Addis

ABSTRACT

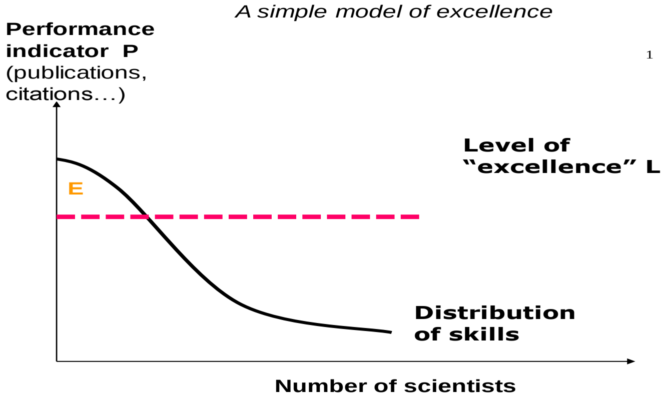
Why are there so few women among the scientists who are considered "excellent"? Excellence is not a naturally given variable. It is produced by a social process of screening for specific abilities, by using specific standards to measure them. This process may be biased. Here I build a simple graphic model for the production of excellence and examine its properties under different assumptions concerning what abilities are relevant and what are adequate standards to measure them. I conclude that the definition of excellence and its creation are the result of subjective as well as objective factors, and therefore contested terrain. Who is considered "excellent" depends on gender relations in the scientific community and in society at large. To complete the argument, this paper introduces an additional model of scientific engagement that explains why more women than men may end up outside the scientific network

In 2003, when the word "excellence" began to surface more and more often in the language of research and academia , I was invited to participate in a workshop on "Minimizing gender bias in the definition and measurement of scientific excellence", and was asked to write the introduction to the publication of the proceedings. In order to provide a unifying structure for the wealth of different insights the workshop produced I built a graphic model of what "excellence" might be. The report, titled "Gender and Excellence in the Making" was published in 2004 by the as a Directorate for Research of the European Commission publication and my model appears as a set of boxed graphs within the more traditional summaries of the contributions to the book.

The graphs apparently met with some success, as I was invited to present them at other conferences on the theme of gender and excellence. This paper contains a revised and hopefully clearer version of these graphs. Moreover, I connect them explicitly with my contribution to the "Gender and Excellence in the Making" report, i.e. the "honor" hypothesis about why women may be less integrated in scientific networks than men with comparable qualifications.

"Excellence" to many scientists, both male and female, appears as a fairly simple concept, that presumably corresponds to an objective and easily measurable variable. It works approximately like this: we assume talent is distributed among people so that some have more, some less, according to some statistical distribution. Those who reach a given level, established by the scientific community, and apply it, so that it can be measured more or less objectively, to create an "excellence" ranking. They receive salaries and resources to do research accordingly, as well as positions of responsibility, and various awards and honors, up to the Nobel Prize, given to the scientists of top rank.This scheme is illustrated in graph 1:

Graph 1.
A simple model of excellence



The number of people engaged in science is measured on the horizontal axis, while performance is measured by an indicator, such as publications, or their impact, or the University where they obtained their degree, and so on, as measured by some indicator acceptable to the scientific community.

The curved line represents the distribution of the skills which are relevant to being a good scientist.

The straight line represents a given level of the indicator variable L.

If PS is the value of the indicator P associated to a given scientist, S, and if LP is the threshold level of the indicator P agreed by the scientific community as that threshold of excellence, then when

PS > LP ==> ES

i.e. when the scientist S is above the "excellence" threshold L then that scientist is considered "excellent", ES. The area E under the curve represents the number of publications by excellent scientists, i.e. the "excellence" produced.

If this simple scheme were correct, the only remaining issue would be to find good indicators and agree on their merit as well as the relative weight of different indicators P (publications versus impact factor versus university etc.). Then we would be able to compare and rank scientists, and decide who is and who is not in the "excellent" range.

If this model is correct, why is it that we find a larger proportion of men rather than of women among the excellent scholars?

In similar studies, for instance when examining the labour market to understand why few women achieve top jobs with the highest earnings and decision making authority, economists use the demand and supply framework to provide an explanation (e.g. Sonnert and Holton 1996). A supply side explanation claims that the supply of women who have acquired the necessary level of competence and seniority is too small. A demand side explanation claims that demand for certain skills is lower when those skills are embodied in a woman than when they are embodied in a man.

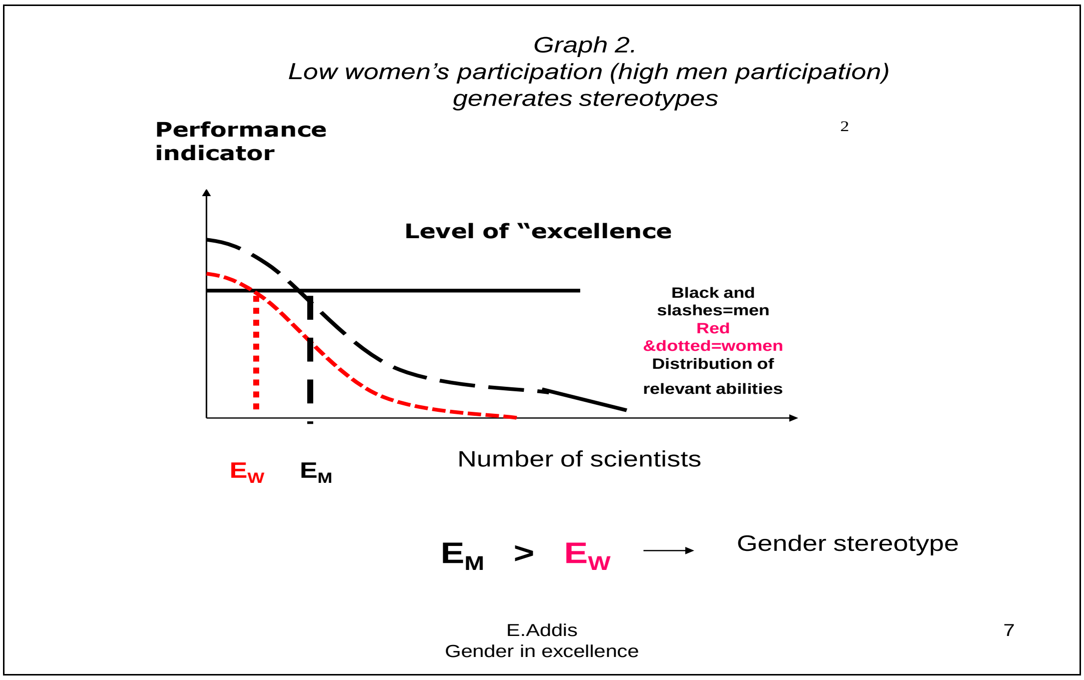
Prof. Larry Summers provided a **supply** side explanation as to why there are few academic women in some fields, that caused him substancial trouble. Most feminist, on the other hand, would choose the **demand** side explanation, i.e. that there is some form of discrimination against women, by male and maybe also by female employers.

This "supply versus demand" debate has been going on for at least thirty years, and by now it looks like a conversation among people speaking different languages. In many cases, and particularly in the case of excellent scientists, neither the demand nor the supply explanation appears to be entirely satisfactory.

There is evidence (Carabelli, Parisi and Rosselli, 1994) that men and women have different perceptions concerning this matter. Academic men claim they do not discriminate against women. Women, on the other hand, think that they are discriminated against. Men would say that they are scientists entirely devoted to the mission of advancing knowledge of their chosen field. If there were women with the skills to do excellent work, why would they not want them in their research teams, for the good of science and of society? Yet, women can present objective measures to show that their careers are not what they should be, given the abilities demonstrated by their grades and other standardized tests. Grades and tests are good predictors of labour market outcomes only within each sex (Benbow and Arjmand, 1990). Why is it so?

The debate shows the widespread perception among women and men scientists that women scientist encounter specific problems either in attaining excellence, in obtaining recognition for their excellence when they do achieve it, or both. Achieving excellence and obtaining recognition, are distinct. They are however connected by acknowledging the existence and the role of formal and informal professional networks, which include potentially excellent scholars and provide enough information to enable some of them to become excellent. (Latour 2005)

Some better explanation is needed: a closer look at the selection mechanisms and at the role of networks may provide some clue. Let us examine the model assuming that there are people of two sexes, equal number of men and women, and that they have the same distribution of skills. The model in graph 1 would apply. If, however, we assume that, for whatever reasons (to be examined later) females participate less in scientific endeavours, the situation changes as described by graph 2.



For purely statistical reasons, even if abilities relevant to the production of excellent scientific work are distributed equally among the sexes, a low F/M ratio in the population of scientists produces a low F/M ratio among the "excellent" scholars, implying that the number of "excellent" men, $E_{M,}$ is larger than the number of "excellent" women E_W.

If F/M < 1 =====> $E_M > E_W$

Notice however that, if it is customary to find more men than women among excellent scholars, then masculinity is apt to become a sign of excellence. The stereotype that men are better than women at doing science is born.

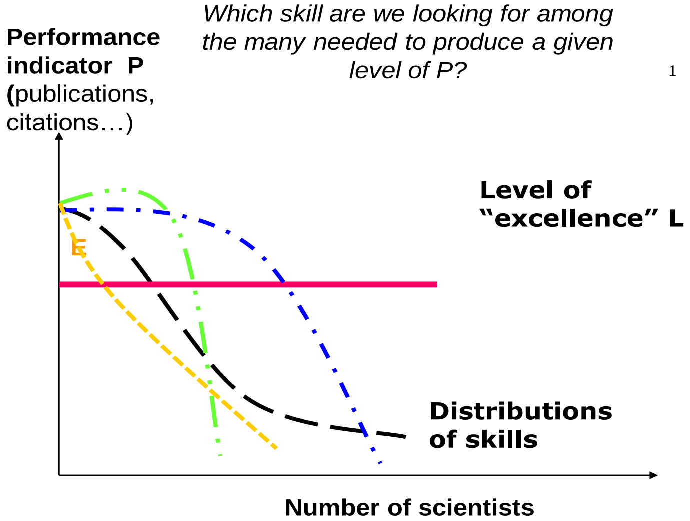
We know from economic theory that collecting information is costly. If being a man increases the likelihood of being excellent, rational employers saving on information costs select men for position which requires excellence unless women with overwhelming higher qualifications are available. Between two people who are alike in any other respect, the male will be chosen. A cycle of statistical discrimination starts (Thurow, 1975). even if abilities are distributed equally among the sexes. Women who are devoting themselves to science, and have the same potential as their male colleagues, do not get the same opportunities to develop this potential.

Next we may ask: what skills are measured by the curve in graphs 1 and 2? Presumably abilities relevant to producing the indicator of excellence P , (most often publications, but other indicators as impact factor, Universities attended can be measured by P). Each different P is the result of the application of more than one skills. For instance, take publications: many characteristics of a human being contribute to producing publishable work: ability to write, being numerate, knowledge of the subject matter, original thinking, choosing a hot topic, and so on. Different observers may disagree about which of these characteristics is more relevant. Some believe that originality is more important than precision, others believe it is the other way around. In any case it is important to remember that we can not measure the skills and their distribution directly. All we can measure is the distribution of the performance indicators, and assume that the distribution of the skill is the same as the distribution of the indicator. By definition, a good indicator is one distributed exactly like the skill we are interested in. But there is no guarantee that it is the case or that our assumption that they are the same really holds. As a matter of fact, any given level of the indicator (black line) is obtained using different amounts of different skills (yellow skill, precision, green skill, originality, blue skill, depths of knowledge, and so on), each with its own distribution.

What we are really looking for is the ability to produce new and useful knowledge. If two people, (for example, one male and the other female, or one white, with an entirely western education, and the other from a less highly regarded ethic group and less prestigious educational system) lead very different lives, the same number of publications may indicate very different levels of ability. This calls for using adjustment factors in order to find out who is likely to produce more new and useful knowledge. For example, the indicator "coming from a good school" must be used only with proper caution. It may make sense to use if other indicators are equal, but not as an additional point in favour of one of two very different candidates. Among other things, good schools are able to give e researcher more secretarial help and better computer technicians, which makes it easier to publish even for a not-so-skilled researcher. In the end, the number of publications is an indicator of skill, but we do not t know what skill other than the ability to produce publications. It is still no guarantee that the person who has more publications will produce new and useful knowledge.

Moreover, it is not always clear what is new or what is useful; there is here a very strong subjective element. New for whom? And useful for whom? What are the aims of those who ask the questions, who finance the research, and therefore get to tell the main story? These questions have already been explored in the literature (Strassmann, 1993), and we know that the issue of power, and the link between power and knowledge are at stake. If two researchers have the same amount of published work, which one do we select as most useful, the one who found out how to increase income according to the old income distribution, or the one who found out how to use income to improve the lives of the most poor? This is a subjective judgement, and in the end it cannot be avoided. Even the judgement of originality is problematic: what is new depends on what the judging person already knows, with the paradoxical result that an ignorant judge may find trivial knowledge original.

Graph 4
Which skill are we looking for among the many needed to produce a given level of P?



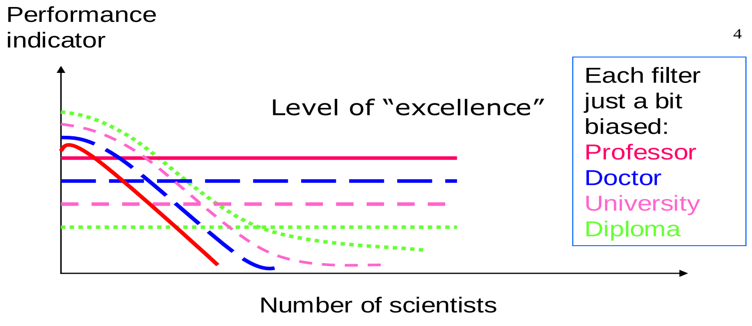
If we accept that there are also subjective elements in judging, we imply that different judges may favour different skills and different indicators. Therefore it matters who the judges are. It is not true that anyone in the postion of judge would make the same choice about who is excellent. Thus the selection is not a neutral process, that always gives the same result, like a chemical reaction.

Further, the criteria will vary by field. The abilities needed to become an excellent mathematician are different from those needed to become an excellent historian. For the former, logic and precision count more, verbal and poetic ability less. Ruthlessness is required form the former, compassion for the latter.

Disciplinary boundaries that define who is admitted in the selection process matter. If the question is about who is an excellent chemist, s/he should be compared with other chemists, not biologists. But the boundaries are not always clear: What about biochemists? What are the proper boundaries of a discipline and who decides what these boundaries are? In economics, for example, critics have noted that the borderline tends to be defined by the methodology rather than by the object of study (Nelson, 1996), and that crossing borders with some disciplines (physics) is approved while intermingling with other disciplines (sociology) is frowned upon. Most often however borderline work is discounted in both disciplines.The history of a discipline is generally not counted as history or as proper part of the discipline. Interdisciplinary work, which some consider more and more vital to increasing knowledge in this era of specialization, risks to become less visible because, in the quest for excellence, it falls in the cracks between disciplines if, for example, only the impact factor of one discipline is counted. New approaches and new fields or subfields may find greater difficulties in being classified as “excellence” exactly because of their diversity and novelty, while the repetition of consensus truth will often be able to link to a web of existing quotation that will produce citations.The quest for excellence, in this case, may lead to conservatism and narrow mindedness, rather than to better science.

Cumulative effects are also at work. The application of the standard of excellence (represented as the horizontal line) is the last step of a series of screening processes that applies increasingly restrictive standards. If each application of the standard is biased in a small way against particular characteristics, this will add up to a large bias against individuals who have them for each subsequent step, so that in the end there will be few of them.

Graph 4. Cumulative effects of a small bias



So, if there is a small gender bias, it is going to snowball. Gender stereotyping, once established, it is going to seep through.

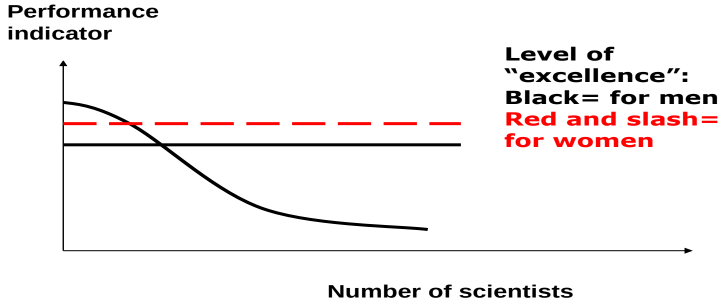
Let’ turn now on the nature of the threshold, represented in the graph by the horizontal line. Is such standard the same for everybody, or, may it set in such a way as to allow for gender bias?

If there is the perception, corresponding or not to reality, that women are not as good as men, then the standard they will have to possess in order to be deemed "excellent" will be higher than the standard men must possess. In her work (Foschi, 2005) and in her contribution to “Gender and excellence in the making”, Martha Foschi provides experimental evidence that in order to be noticed, both by men and by women, women must make more of an effort than men. Additional experimental evidence is provided by Goldin and Rouse (2000) in the well known article according to which in which performance behind a screen increased the chances of women violinists to be selected to become part of an orchestra. If women must pass a higher standard in order to be counted as "excellent", the result is again that

$E_M > E_W,$

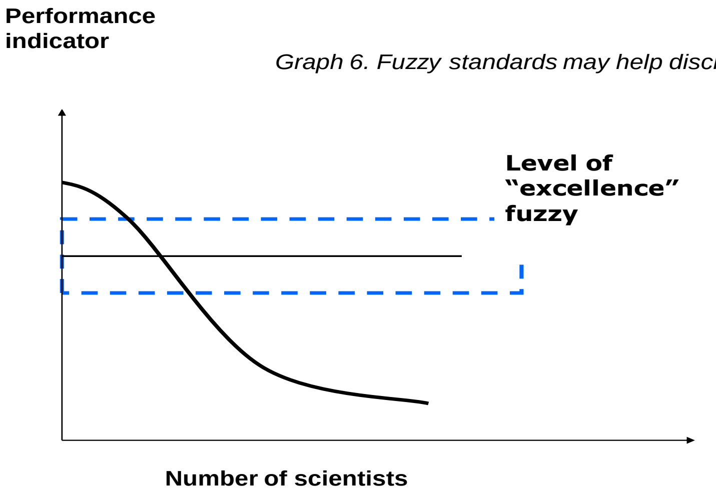
with the risk of further gender stereotyping. The situation appears, as shown in graph 5 below.

Graph 5. Double standards



This is a visual reprentation of the the “double standard” applied to people of different sex. It is easier to apply a double standard if the standard is fuzzy, i.e. unclear to those in charge of the judgement and unclear to those who should pass the standard. The horizontal dotted line in graph 6 represents fuzzy standards, i.e. standards that may be or appear higher for women than for men, reducing the number of women among those considered "excellent"

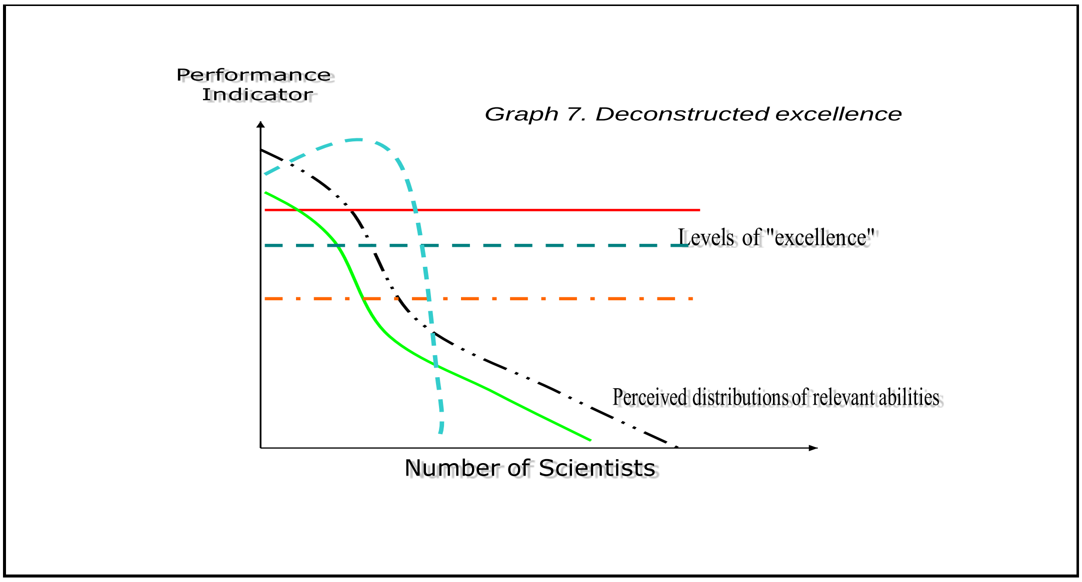
Graph 6. Fuzzy standards may help discrimination



We are now able to summarize the factors that affect processes of selection on the way to excellence, that may contribute to the scarcity of women at the top.

a) the small number of women among scientists may affect the F/M proportion among the “excellent” scholars creating an unfavourable stereotype
b) the indicators of performance chosen by the judges intended to measure ability, but in fact correspond to many abilities. What is the most relevant ability, and how it is distributed, is a matter of perception. Therefore different “gatekeepers” may judge differently. Since there is a tendence for judges to prefer people similar to themselves, once gatekeepers who have one kind of ability are established, they are likely to favour people with that ability, thus perpetuating the bias.
c) The boundaries of the discipline matter: borderline, interdisciplinary, and very original work may be less visible because it falls between boundaries
d) The repeated application of a small biased standard creates a large bias at the end of many cycles
e) Once the unfavourable stereotype is established, this is likely to lead to double standards. A person who belongs to the group stereotyped as less competent will have to perform better in order to be recognized as "excellent". Hence, women must be better than men to achieve the same rank.
f) Such setting of double standards is easier if the standards are fuzzy. If neither the judges nor the judged know exactly what is it that is being measured and how, it is easier to set higher standards for one group, such as women.
g) Moreover, if some people are less optimistic, rather than less competent, they may avoid competition believing that the standard is too high for them. There is evidence that women as less risk-takers than men.

If we put together all these possible sources of bias in the process of production of excellence, the model is better illustrated by the complicated model of graph 7 rather than the neat, simple one.



Clearly, different people may be judged "excellent", depending on the opinions, of the judges and the criteria they use. The definition and creation of excellence becomes a contested terrain, and its attribution may well be related to gender relations in the scientific community and in society at large. When finding an agreement on the criteria to define excellence scientists are not discovering the “true” model of excellence, as defined in graph one. They are rather building what will become the operationally used model of excellence.
It seems a minor difference but it is not: it is the basis upon which we can hope to change the operationally used model of excellence, in such a way that more women can rise to excellent status without having to perform exactly how men in the past have done.

A different model.

I now introduce a second model, that, without any mechanisms of discrimination examined above, results in more men than women choosing science and therefore then, according to the mechanisms examined before, more men then women ending up among those considered “excellent scholars”. This model is a novel and important complement of the one above.

Why women participate less <u>and men more</u> to the scientific enterprise? Because of two sets of reasons, the first set due to women’behavior, the second set due to man’s behavior.

a) women choose to perform domestic activities, because they think they are more important than doing science. As economists would say, the utility they derive from activities other than doing science is greater.
b) men like to play at science more because of the “game” aspects of the activity: they love the contest. If science is seen as a fight to establish one’s results and reputation, the utility they men from doing science is greater
c)

Usually reasons under a) are considered the main if not only culprit: Women think that they have better things to do with their time, attending to the children, housekeeping, tending the networks of relationships and to the needs of the the elderly. The problem of women in science is treated as a problem of conciliation of work and family duties, and the solution id offering more services as an alternative.

These reasons certainly play an important role, and I do not wish to understate it. In economically developed countries, though, where fertility has declined and the public and private sector provide schools, take-home-meals and nursing homes, this set of reasons is becoming less important.

Rather, I want to emphasize why men devote so much time and energy to scientific endeavors, sometimes even to the detriment of their children, of their relationships, of leisurely activities, even of their health.

My hypothesis is that they do it because they like the “game”, an activity where they may win or loose. If this hypotheses is correct it also helps to explain why they may wish to exclude women from this activity.
Scientific activity is rewarding well beyond the monetary compensations that most scientist of either sex receive. The satisfaction of having added one’s bit to human knowledge, the hope that it may help somebody to lead a better life, are rewards for men and women. However there is also a competitive part, i.e. the run to be the best scientist. As there is a different enjoyment in just running in the park and competing in the Olympics, there is a difference between enjoying science for its inherent pleasures or enjoying it because it is a game, with rules and a scoring mechanisms capable that produce ranking.

Any encounter between two scientists is also an engagement where each explores the credentials of the other (where did s/he get the Ph.D., whom does s/he knows, what s/he published) to check the relative standing in the profession. Particularly in the U.S., but now increasingly in Europe as well, seminar at times erupt in serious battles where ideas are challenged and defeated. The result of such engagement is exchange of information. Information flows through the formal and informal scientific exchange, both about interesting research and about social relations in the scientific community.

Game theorists analyse games by means of pay-off matrices. Are men and women playing the same game? Do they have the same pay-off matrix ? Suppose there is a perception, as expressed in the ancient idea of honor, that men gain honour by fighting other men, not only if they win but even if they lose honorably; but a that a man should always win when fighting a woman, who is a member of the weaker sex. Then, losing to a woman is a major loss of honour. If this is the case, the pay-off matrix of a man engaging a woman appears as follows:

Payoff matrix in an "honor" system

	Competition (intellectual) with a woman	Competition (intellectual) with a man
Gains from winning	SMALL	LARGE
Losses	LARGE	SMALL

Engagement in scientific conversation with a woman scientist pays less for a man scientist than engagement with another man. If this is the case, and the man can choose whether to engage with one of two scientists of different sex and of the same perceived rank, he will rationally choose to talk to a man, rather than a to woman, because the chances of gaining status are greater.

If this is the true, the result will be that women don't get the same flow of information, they remain "out of the loop". They may feel that their ideas are not worth challenging. Their sense of self develops differently than that of their male colleagues (Aberson CL, M Healy, and V Romero (2000)).. They are at a disadvantage that may quickly become cumulative.

The matrix above describes the the payments to the choice, under the assumption that the game is played by a man scientist. He is the one who has to choose whith whom to engage. We may ask the question whether the matrix of payments would be the same when the game is played by a woman scientist. She may or may not perceive the payment as above, or elese, she may prefer to engage with a woman scientist. This is an empirical question, which may be worth esploring. However, as long as the matrix above represents the the pay-offs of some actors, the result that women scientists are at a disadvantage holds.

CONCLUSIONS AND POLICY GUIDELINES

The main conclusion of the first model in this paper is that when a gatekeeper is assessing someone's excellence, s/he is not making the only possible judgement of excellence, but rather making a choice of a specific set of possible criteria, and applying these to people whose life may have been very different because they were born female or male. Notions of excellence are often considered to be value while they are not, nor is possible to make gender neutral judgements about excellence. Thus gatekeepers are exercising a discretionary power that is not gender neutral.

The main result of the second model is that discrimination against women may derive by the simple fact that men and women have different tastes about fighting, not different abilities.

Reality may well have some relation to both models.

Excellence is not given as a fact, but rather a set of practices used to organize the scientific community. The model presented in this papers were born when trying to give visual clarity to the analysis of these practices by the research group that produced the report "Gender and Excellence in the making" (Addis and Brouns, 2004) which I was in charge of editing. The analysis of these practices suggests that many women capable of producing useful and even excellent scientific and technical knowledge are kept out from the top levels, where strategies of research are decided, while men with lesser abilities may may reach top positions. The selection process for excellence is such that it can be rebuilt according with more consideration to gender differences in life cycles and to different relational attitudes of men and women. If the scientific community wishes to increase the number of women scientists at all levels, because this will lead to more progress in science, the entire process – indicators used, selection of gatekeepers, procedures - should be revised accordingly.

Bibliography

Aberson Cristopher L, Michael Healy, and Victoria Romero (2000). Ingroup bias and self-esteem: A meta-analysis. *Personality and Social Psychology Review* 4:157-173.

Addis Elisabetta, and Margo Brouns (2004) "Gender and Excellence in the Making" Directorate General for Research, Science and Society series, EUR 21222, Bruxelles, 2004.

Blank RM (1991). The effects of double-blind versus single-blind reviewing: Experimental evidence from the *American Economic Review. American Economic Review* 81:1041-1067

Benbow Camilla and Olya Arjmand (1990). Predictors of high academic achievement in mathematically talented students: A longitudinal study. *Journal of Educational Psychology* 82: 430-441.

England P, P Allison, S Li, N Mark, J Thompson, M Budig, and H Sun (2004). *Why Are Some Academic Fields Tipping Toward Female? The Sex Composition of US Fields of Doctoral Degree Receipt, 1971-1998,* http://www.stanford.edu/dept/soc/people/faculty/england/Tipping.pdf.

Ferber, Marianne A., and Michael Brun "Does your legislator sex matter?" Policy Matters, 4, 1, Autumn 2006: 11-15.

Fletcher J (2001). *Disappearing Acts. Gender, Power, and Relational Practice at Work.* Cambridge, MA: MIT Press.

Foschi, M. "Gender and Double Standards in Competence Assessment." Pp. 140-48 in V. Maione (ed.), Gender Equality in Higher Education. Milano: Franco Angeli, 2005.

Folbre, N.. 1993. "How Does She Know? Feminist Theories of Gender Bias in Economics" History of Political Economy 25(1): 167-184.

Ginther D (2006). The economics of gender differences in employment outcomes in academia. In *Biological, Social, and Organizational Components of Success for Women in Academic Science and Engineering*. Washington, DC: National Academies Press (In Press).

Goldin C (2002). *A pollution theory of discrimination: Male and female differences in occupations and earnings* (Working Paper 8985). Cambridge, MA: National Bureau of Economic ResearchGoldin C, LF Katz, and I Kuziemko (2006). *The Homecoming of American College Women: The Reversal of the College Gender Gap* (NBER Working Paper 12139). Cambridge, MA: National Bureau of Economic Research.

. C. Goldin and C. Rouse (2000). Orchestrating impartiality: The impact of "blind" auditions on female musicians. *American Economics Review* 90:715-741.

Heilman ME, AS Wallen, D Fuchs, and MM Tamkins (2004). Penalties for success: Reactions to women who succeed at male gender-typed tasks. *Journal of Applied Psychology* 89(3):416-427.

Herlitz A, LG Nilsson, and L Baeckman (1997). Gender differences in episodic memory. *Memory and Cognition* 25:801-811.

Jacobs J and S Winslow (2004). The academic life course: Time pressures and gender inequality. *Community, Work and Family* 7(2):143-161.

Krefting LA (2003). Intertwined discourses of merit and gender: Evidence from academic employment in the USA. *Gender, Work, and Organization* 10(2):260-278.

Klamer, A.,D. McCloskey and R. Solow (eds) The Consequences of Economic Rhetoric New York 1988.

Nelson, J. A. Feminism, Objectivity and Economics Routledge and Kegan Paul, New York,

Latour, B *Reassembling the social: an introduction to Actor-network theory*, Oxford ; New York, Oxford: University Press, 2005. ISBN 9780199256044

Rolison DR (2003). Can title IX do for women in science and engineering what it has done for women in sports? *American Physical Society News Online* 12(5):8.

Rosser SV (2004). *The Science Glass Ceiling.* New York: Routledge.

Skolnik M (2000). Does counting publications provide any useful information about academic performance? *Teacher Education Quarterly* 27(2):15-25.

Strassman D. "The Stories of Economics and the Power of the Storyteller." History of Political Economy 25(1), 1993, pp.145-167

Sonnert G and G Holton (1996). Career patterns of women and men in the sciences. *American Scientist* 84:63-71.

Trix F and C Psenka (2003). Exploring the color of glass: Letters of recommendation for female and male medical faculty. *Discourse and Society* 14(2):191-220

Wennerås C and A Wold (1997). Nepotism and sexism in peer-review. *Nature* 387:341-343

Xie Y and KA Shaumann (2003). *Women in Science: Career Processes and Outcomes*. Cambridge, MA: Harvard University Press.

Zimmer L (1998). Tokenism and women in the workplace: The limits of gender-neutral theory. *Social Problems* 35(1):64-7