December 2, 2012 meeting about meta-research

Participants:

- **Dario Amodei (Stanford University, School of Medicine)**
- **Kristen Fortney (Stanford University, School of Medicine)**
- **Jacob Steinhardt (Stanford University, Computer Science/AI)**
- **Yoni Donner (Google; Stanford University, Computer Science/Comp Bio)**
- **Jade Q Wang (Meteor, formerly Northwestern U, Neuroscience)**
- **Paul Christiano (UC Berkeley, Theoretical Computer Science)**
- **Nick Conley (HGST, formerly Stanford University Chemistry/Biology)**
- **David Dalrymple (Nemaload Project, formerly MIT)**
- **Holden Karnofsky (GiveWell)**

Note: This set of notes was compiled by GiveWell and gives an overview of the major points made by the conversation participants

General Discussion of Academia and Science

Scientific equipment

There is a market inefficiency in the area of scientific equipment arising from the fact that governments are the effective buyers of scientific equipment and are to some degree willing to pay the market rate independently of the value of the equipment, so that there aren't strong incentives for science equipment manufacturers to make cheap and useful equipment.

Scientists do want better equipment that serves functions that they're familiar with, and this gives rise to healthy competition between suppliers of some commonly used types of equipment such as mass spectrometers. However, scientists sometimes don't know that certain types of equipment or functionality are even possible, and so don't give suppliers an incentive to develop them.

Difficulty with matching theory with desired applications

Because theoretical research takes a long time to progress toward the point of yielding applications, it is difficult to predict whether pursuing a given set of theoretical questions will be useful to applied scientists when the research finally comes to fruition. What applied scientists need now may be very different from what they need in 5-10 years.

Theoretical research sometimes results in unexpected applications. For example, the machine learning research on Bayesian networks in the 1980's has recently been applied to the study of biological networks. (However, some researchers have found that Bayesian networks do only a little bit better than simple linear models in their areas of research.)

Gulf between theoretical and applied researchers

In some cases there are techniques that theorists have developed that applied scientists would find useful but don't know about. It might be good for there to be people whose job is to interface with both the applied scientists about their needs, and the theoretical scientists about what they can do and communicate these things to the two groups.

Interdisciplinary activities

In some cases, techniques that have been developed in scientific field A can be used to solve a problem in scientific field B. Often the researchers in field B don't know enough about field A to be able to do this. This is especially the case when field A and field B are sociologically distant, as is the case with biology and math.

It could be that trying to find connections between field A and field B would not be a high leverage activity because there are so many possible combinations of things from the two fields to look at.

It would be valuable to have scientists switch fields to cross fertilize, and scientists who did this would, on average, produce more than they otherwise would, but the structure of academia discourages switching fields, and doing so constitutes a major career risk. Perhaps funding postdocs for recent PhDs who want to switch fields would be a good philanthropic opportunity (there already exists such a program for people switching from physics to biology, but more such programs might be needed).

The need for greater interdisciplinary work is acknowledged in academia and there have been many efforts to promote it, but despite these efforts the need does not seem to have been fully addressed.

Transitioning from and to academia

If a researcher leaves academia for several years, works in industry or a start up, and returns, the academic world generally considers this to be equivalent to the researcher not working for those years, rather than considering the researchers' output during that time to have value.

If an academic is going to try working at a start-up, it's often best for his or her career if he or she does it in graduate school rather than later. Leaving academia after getting a PhD typically severely limits one's prospects in academia.

However, there are some exceptions, eg:

- Stanford is unusually sympathetic to science graduate students who want to work at a start up while in graduate school.

- The robotics program at MIT seems to look upon industrial experience with robots favorably when evaluating graduate student applications.

Overly strong emphasis on metrics in peer review

Frequently, the most compelling aspects of a research project will be difficult to quantify, and one can only make an intuitive argument for their value. As such, sometimes the emphasis on metrics in the peer review process doesn't capture the value of a project.

Peer review in theoretical physics and computer science

In computer science, research articles are usually published in conference proceedings rather than journals. In theoretical physics, they are simply uploaded to the arXiv (a database for preprints) and often not published in journals at all.

Compartmentalization

Academia is very compartmentalized, which cuts down on opportunities for collaboration.

Disconnect between fundraising and value

The professors who get the most funding tend to be the professors who are best at selling their projects rather than the professors who are doing the most important work.

Negative feelings toward industry

Within parts of academia there is a negative feeling toward industry and toward academics who leave academia or do something outside of academia. Stanford is exceptional among universities in valuing work outside of academia.

Mismanagement of human capital

In some experimental sciences, particularly in experimental biology, there is a hierarchical culture in which all of the members of a lab have to follow the orders of the principal investigator. This selects for people who are obedient and focused on following rules rather than on efficiency. It also drives highly talented young scientists away from academia.

If a member of a lab is able to work unusually efficiently, he or she is apt to be assigned additional work rather than being given free time to do his or her own research.

The cost of graduate students to a principal investigator is typically very small, so he or she will often see their labor as less costly than equipment and software, and so assign them large amounts of work that could be automated with small expenditures. But the opportunity cost of their labor is very high, because time spent on automatable tasks could be spent on learning and career development. The principal investigators don't internalize this opportunity cost, and so are frequently insensitive to it.

Subject specific issues

Funding for aging research

The National Institutes of Health (NIH) has a budget of 1 billion dollars for aging research. Half of it goes to Alzheimer's disease research. About a quarter goes to geriatric care (e.g. research on how to improve the quality of nursing home care). The other quarter is devoted to miscellaneous projects in the study of aging. Most of it is not focused on slowing or preventing aging in humans. Few university departments work in this area, and there is relatively little interest in this area even though success would have very high value.

If the NIH adopted aging research as a focus then scientists would attempt to frame the projects that they were working on as being relevant to aging research, just as they currently frame their work as relevant to cancer. Thus, the consequences of a shift might be complex and hard to predict; it may be better to reallocate aging research to better aging research, than try to increase the aging budget overall.

The Glenn Center for the Biology of Human Aging Research

The Glenn Center for the Biology of Human Aging Research is one organization working on the cause of slowing aging in humans.

Types of aging research

Senescent cells are cells that have stopped dividing. There is evidence that they are poisonous to the body and cause cancer. Systematically killing these cells could potentially prevent aging. However, only two labs are working on this.

There has been work on sequencing genomes of people of age 110 or older and looking for common genes across them with a view toward isolating genetic variants that correlate with extreme longevity. However, there are only 22 such people currently alive in the United States. One could broaden the scope of the study to people over 100 — there are thousands of such people in the United States. However, there is not enough money to fund such research. It may be premature to do this project, because the cost of sequencing a genome is currently about $4000 and is expected to drop over time. In the immediate future, a centenarian DNA sample collection initiative could be of very high value. Collecting ~1000 centenarian DNA samples would take 1-2 years, at which time the cost of sequencing will likely have dropped substantially.

It would be good for research if more genomes were made public (of any type – centenarian, normal or disease), but there are privacy concerns.

Howard Hughes Medical Institute (HHMI)

HHMI is a strong biomedical research organization that in some ways represents a system parallel to academia, but which allows for more risk taking. Publications from HHMI labs (e.g. Janelia Farms) tend to be fewer in number but have more citations than average, suggesting that HHMI's environment encourages risk-taking. The HHMI Investigator Program is good — it finds Principal Investigators who have great track records and gives them freedom. It would be good for philanthropists to fund similar programs, and in fields outside of biomedical research as well. However, there is also some concern that the HHMI system, by providing freedom and lots of funding, may remove accountability.

Machine learning

In the field of machine learning:

- Almost all papers are in the public domain (e.g., available on researchers' websites).
- Code is not always shared, but there is a cultural incentive to share code.
- Engineering details concerning implementation are not always published. Researchers tend to write about the theoretical framework that they're using rather than the engineering details.
- There is a great deal of interdisciplinary collaboration.
- It would be good if there were larger data sets available. There's little incentive for individual researchers to compile large data sets. Processing large data sets requires good computing infrastructure, so there is a need for better computing infrastructure.
- It's hard to gauge general progress, because researchers design algorithms for various specific types of data sets.
- For many benchmark datasets the test data is publicly available. It's difficult to exert discipline to not look at the test data, and so machine-learning researchers sometimes (consciously or unconsciously) overfit their models to the test data. It would be desirable for test data to be inaccessible to researchers until they commit to publicly releasing the test results.
- There's an overemphasis on public relations and researchers sometimes oversell their products.
- Machine learning has applications to many fields. Software companies are typically users rather than developers of machine learning. A high leverage area might be that of making machine-learning research easier for software companies to use.

DARPA Grand Challenge

Defense Advanced Research Projects Agency (DARPA) put on a series of contests in robotics called the DARPA Grand Challenge where contestants write programs for a specified robot to do a particular task, and are awarded a very large ($100 million) prize if they submit the best entry. This contest's winning entry ultimately led to Google's self-driving car. Since then, people have improved on the original entry, but the impetus was

DARPA. The invention of the self-driving car might ultimately provide billions or even hundreds of billions of dollars in value, making the prize a potentially very cost-effective investment.

High Resolution MRI

Current magnetic resonance imaging (MRI) technology is sufficiently high resolution for doctors' use, but not high enough for researchers to do the projects that they would like to do. The resolution is typically 1 millimeter. Recently, researchers developed MRI that can scan at 1-micron (0.001 millimeter) resolution, but it can only scan a single square millimeter of the brain, and it would be significantly more useful to have a MRI that could scan the whole brain at 1-micron resolution.

Cancer budget reallocation

Many projects that are ostensibly cancer research are not actually cancer research, but relate to general cell biology in a way that might (often vaguely) have a bearing on cancer. Cancer thus in some ways provides an "excuse" to study general cell processes, but it's not clear if this is really the optimal way to fund basic research. It also makes funding statistics potentially misleading.

In attempting to increase funding for a particular area of biology, one should be aware of this "fungibility" effect.

To gain a better understanding of the situation, one could try to make a preliminary assessment of how much money allocated to cancer research is actually going to cancer research by combing through the NIH proposals for cancer research funding, which are online.

Psychology

In the field of psychology, studies are often not replicated. The studies often have too few subjects for the results to be statistically significant. Even if they're ostensibly statistically significant, the tests for significance implicitly assume a large sample size. The subjects of the studies are also usually undergraduates in North America, and so may not be representative.

One reason why the sample size is small in psychology is that subjects have historically needed to be paid individually. However, there are emerging ideas for ways to collect data for psychology studies very cheaply, using online data sources or clever hacks (e.g. experimenting with diet using voluntary sign ups in Google's kitchen).

Diet and Exercise

There seems to be a paucity of research on nutrition and exercise. This may be because there is a lack of demand from corporations (which may not have much to gain from such research) and government health care (which doesn't focus on nutrition and exercise).

Computational neuroscience

The field of computational neuroscience mostly includes programmers rather than neuroscientists. They have an aversion to experimental research, and tend to believe that brains are wired randomly. Data sharing is very poor in computational neuroscience.

Bioinformatics

In bioinformatics, data sharing is much better than in computational neuroscience because it's required by journals.