

# Rough draft notes by Paul Christiano from 2013

## Introduction

Suppose that I am making decisions about my career and am interested in supporting project X. I could pursue a career that directly contributes to project X, but sometimes it may be more effective to pursue a job which uses my skills and pays well, and to use the money to pay others to work on X.

The natural approach to the decision (at least for people with a certain mindset) is to try to estimate the counterfactually adjusted impact of dollars for project X, the counterfactually adjusted impact of talent for project X, and my earning potential, and then to chug through an expected value calculation. Phrased in this way the answer seems to depend on the notion of *replaceability* [LINK]: how likely would someone else be to do this job if I didn't? It also involves a host of other factors, including understanding my career prospects, the kind of work that needs to be done in project X and the impact of that work, the cost of hiring a comparable researcher, and so on.

However, we can partially avoid asking difficult-to-answer questions about our earning potential or the cost of hiring replacements who don't care about X, because our uncertainty about these questions are correlated. I think that by making more direct comparisons we can not only save effort but also get a significantly better answer. [LINK]

In particular, I think that the answer to our original question is roughly the same as the answer to the simpler question:

" Compared to the marginal person who would work on X if the wages were raised, do I have comparative advantage in working on X or in earning? "

This allows us to make our decision by comparing apples to apples; we now need to think about how well-suited we are to a line of work, compared to the other people considering that line of work. This is the sort of comparison which a self-interested career-seeker must also address, and which is quite close to our experience (and moreover the world is structured to make this question easy; we have price signals, discussion with advisors, and interaction directly with the relevant people).

Overall, the tentative result of this analysis is that direct work seems to be undervalued

within the effective altruist community. The main reason for skepticism regarding earning to give is that if I can pursue a very lucrative career, then someone similarly competent would also have similarly lucrative opportunities. Thus if all else is really equal (and we don't have comparative advantage in the lucrative job) then we would have to pay a prohibitively large salary to induce them to take the direct work. So in order for the opportunity to be favorable, we must in fact be exploiting comparative advantage (e.g. because the direct work isn't exploiting our most valuable skills).

In fact I think the argument we are making here applies somewhat more broadly than to money. For other kinds of social capital and influence we can run a similar argument, which is somewhat less convincing primarily because transferring social capital is more difficult. However, I would still accept this as a rough guiding principle in many other cases.

## **The picture**

### **A simple model**

Here is a very basic model of the situation:

In the status quo, I am volunteering working on X, creating value V. Maybe I spend some money on myself, but that will be the same across all scenarios so I'll neglect it and assume I'm living off exactly enough savings. Also, I might be being paid to do X rather than volunteering, but for simplicity let's assume that donating to cover my salary is my preferred use of money (which makes the situations equivalent).. None of these manipulations have any effect on the result (the diligent reader can convince themselves).

I am considering quitting X, and taking on an unrelated job Y which pays a good salary S but creates little value. If I did so, I would then take the wages S and use them to pay someone else, say Alice, to work on X. (In fact I might also spend a bit more or less on myself, to compensate myself for whatever change in quality of life I experience---I want to consider the change that has no effect on my own welfare.) Alice will then stop working on an unrelated job Z and start working on X, creating some amount of value. Of course the amount I have to pay Alice is roughly such that Alice is indifferent between taking job Z (which offers a salary of S') and working on X. Maybe I make enough to pay several people to switch, or maybe I make enough to pay less than one person to switch. Maybe people are of inhomogeneous quality, etc. The people I replace

will create value  $V'$ .

Prima facie, we expect that the wages we would have to offer Alice are roughly  $S'$ . In this case, I would prefer quit X if and only if  $(V' / S') > (V / S)$ , i.e. if and only if Alice has comparative advantage in working on X. Of course, the devil is in the details, which we'll get to shortly.

## **The big considerations**

Here are two considerations that bear on whether your comparative advantage is lies within X. These seem to me to be the two biggest considerations, and (unfortunately) to pull in different directions:

1. Most activities in the world don't contribute to X, and so entropy is on the side of earning to give. If you were good at one random thing, it is much more likely to be something someone will pay you to do than to be something that contributes to what you feel is the most important problem. This depends in large part on how broad you feel the class of "maximally important" projects is, and on how transferrable skills are between projects. If you think a very small set of activities is really important, and that your skills are very localized to a particular project, then this consideration can easily completely dominate. I think the class of nearly maximally important projects is reasonably broad and that skills transfer reasonably well, and so this is a large but not overwhelming consideration. Moreover, this consideration doesn't really apply if you are considering doing something like finance just because it earns a lot, it only applies if you used knowledge of your comparative advantage in the selection process itself.

2. Paying someone to do something they don't care about intrinsically is hard. By virtue of caring about X I can work harder in many unmonitorable ways, exhibit more creativity and flexibility, seek complementarity between the rest of my life and my work, and just generally be much more effective. In practice this effect seems to be quite large, though perhaps not larger than pursuing work that I enjoy.

Fortunately these two considerations can easily factor into our intuitive impression of "do I have comparative advantage in working on X?" We are familiar with what we are good at, and we have a strong reference for people being paid to do things they don't care about. We can learn more about both of these facts by introspection and exploration (for the first) or just looking around the world (for the second).

Note that the first of these considerations should manifest as a comparative advantage

in some particular lucrative career under consideration. So if it's unclear where my comparative advantage is, it seems that consideration (2) will tend to dominate, and so I would tentatively recommend direct work rather than earning to give.

In some sense consideration (2) is a consequence of market imperfection: under similar hypotheses it would be a Pareto improvement for someone in a high-earning career to work a bit harder and receive a bit more money which they could use to pay more. But even with 0 friction the corresponding bargaining problem is too hard, and so at best the hard-worker captures some fraction of the gains from their harder work. So even without friction there is still significant underinvestment, and in the real world the situation is massively worse than that.

## Revisiting the argument for earning to give

A common motivation for earning to give is the observation that it is quite plausible for a motivated and talented young person to expect to earn \$20M over a career in finance or technology [LINK], which appears to be much larger than the cost to employ several talented researchers for their entire careers. So it seems that taking a high-paying career and making donations to organizations that employ researchers is significantly better than working as a researcher myself. Relatedly, it seems that if I work directly then I am displacing someone else who would have been hired, whereas if I earn then I am not replacing anyone and every dollar I contribute is additional [LINK].

The difficulty, as indicated by the above analysis, lies in the following question: if people with my level of competence can in fact expect to make \$20M over a career, why would they seriously consider a job as a researcher making several times less? Although *homo economicus* is not a great approximation, when one sees a deviation this large something is typically off. In this case:

- It is easy to overestimate my own potential lifetime earnings.
- It is easy to underestimate the difficulty of working on X.
- Good researchers work for little money, typically because they love doing research. I have seen many people turn down factor of 2 or larger multipliers on salaries to pursue a career they enjoy. So if you want to entice away such people, you can't just provide money, you need to provide a job they find compelling enough to accept a below-market salary.

So with these arguments in mind, let's return to our simple model.

## Squaring with the model above

There are some conditions under which the argument for earning to give seems to work, and in these cases this is manifest in the comparative advantage picture. These considerations should still be weighed against others, and the comparative advantage picture gives a relatively easy way of understanding how to do the weighing.

1. If your skills aren't relevant to X, there is a good chance you shouldn't be doing X.

2. If other people do X in large part because it makes them feel happy or satisfied, and you don't feel happier doing X than doing other things (say because you get the same warm glow by donating as by doing good directly), then you should pay someone else to do X, so they can enjoy the happiness which is lost on you. For example, if others work as lawyers to non-profits because they like the feeling of doing good, and you would feel just as good by earning and paying some lawyers for non-profits, this is an argument in favor of earning.

3. If other people do X because it's a nice job and you don't care about having a nice job, then you should do some worse job that pays well. For example, it may be that you can earn much more in a job where you work 80 hour weeks, and you could employ many people to do direct work with your salary. But in this case we should be careful to understand the actual asymmetries here: there need to be not one but two of them. First, you need to actually care less about your own welfare than other job-seekers (who are passing up the 80 hour week). Second, X needs to be a job where suffering is less useful than other jobs on the market: you can do many things better by dedicating your life to them fully at the expense of personal welfare, so in order for this to distinguish between earning to give and doing good, personal sacrifices (e.g. working 80 hour weeks) need to be more effective for earning.

4. If other people do X because they care about the success of project X, but they are much happier to accept lower wages than to earn higher wages and donate, then it may be better for you to earn to give (this is more common but less compelling than the other cases). For example, if you are considering becoming a doctor because you care about helping the sick, and other would-be doctors are also willing to accept low wages because they care about helping the sick, but other would-be doctors *aren't* willing to

use their wages to help the sick, then this introduces an asymmetry that cuts in favor of earning and donating.

For this to work, it must be the case that people can be attracted to work on X based on their altruistic inclination, and that those people weren't working on X anyway (for example, this might happen if some popular causes had various flavors, some of which were much more socially impactful than others). The size of the advantage is given by how large a pay cut people are willing to take to work on X. In many cases this will need to be balanced against the fact that (by hypothesis) those people don't actually care very much about X itself and so are unlikely to take a very well-focused approach to working on X unless the job is fairly well-circumscribed.

In each case there is a clear explanation for where your comparative advantage is coming from: (1) doing a job that uses your skills, (2) avoiding a job that doesn't make you happy but makes other people happy, (3) doing a job that involves large sacrifices, if you are unusually willing to make such sacrifices, (4) donating to a cause that other supporters are unwilling to donate to.

## **Details**

### **Who is the marginal person?**

An important question in this picture is who exactly we are substituting for. In the basic picture above, this is the best person (to hire for X) who is not yet working on X. If X is a big project, it can probably be reasonably approximated as the worst person (to hire for X) who is currently working on X.

The nature of this marginal person captures intuitions about whether an organization is "funding constrained" or "talent constrained." For a "funding constrained" organization the marginal hire has comparative advantage in working on X, and so it is more likely to be a good idea to do work elsewhere and provide funding. For a "talent constrained" organization the marginal hire does not have comparative advantage at X, and so it is more likely to be better to work directly for them (rather than forcing them to hire someone who is a poor substitute or demands very high wages). In general, it seems easiest to reason directly in terms of the marginal person, rather than to reason about the intermediate variable of funding-constrained or talent-constrained.

For some projects X, the marginal hire is quite enthusiastic about X, and so the question is often about comparative advantage amongst people who are enthusiastic about X. In this case it is much more likely to be reasonable to earn money that can be used to hire enthusiasts. In a simple picture, where the world is divided into "people who care about X" and "people who don't" the result is likely to be that enough of the "people who care" earn money that they can fund the others, and the people who don't care are never involved with the project except in roles where the enthusiasts lack relevant expertise. In practice things aren't so clear, but a similar situation will occur whenever there are more people who want to work on X (for the sake of accomplishing X) than there are other funders. In this case, some people will need to provide funding instead, and they should be those with a comparative advantage at earning amongst X-enthusiasts.

### **Asking employers**

I opened this post by discussing the bean-counter's approach to the problem of earning vs. doing direct work. There is a somewhat more direct approach: directly ask your would-be employer, "would you prefer I take a job with you, or that I donate \$N a year?" Of course this requires a would-be employer to be fairly familiar with your ability to contribute, perhaps as a result of having worked together in the past, the interview process, or comparisons with peers.

Naturally this is an input into the process I have described so far. You are basically asking "how much do you think you would have to pay to get the same amount of value I could contribute by taking this job?"

Within the process I have described, a more natural alternative would be "What is your marginal hire like? Do they seem more or less well-suited for this job than I am, relative to other options?"

There are a number of advantages to this alternative (even if you can't directly ask this question), and more generally towards thinking this through using some framework like the one I propose rather than deferring to a would-be employer's answer:

\* Your earning potential is highly uncertain, as is the actual quality you can contribute as an employee. These facts are highly correlated with each other, and both you and your would-be employer are highly uncertain about both of them. Unfortunately, both you and the employer only know part of the relevant information. If you estimate independently, your answer is influenced a lot by random noise in the variation of your estimates [LINK]. I think the alternative question better isolates the information that the employer

possesses that you don't.

- \* You can make this comparison relatively easily between your *current* job offer and your *current* opportunity to do altruistic work directly. These are reasonable proxies for the rest of your career, but they are still proxies. Both you and your would-be employer need to reason about the long-term question.

- \* Given uncertainty, I think a would-be employer needs to reason through this problem in something like this way (or, more often, they will rely on some implicit heuristics that may or may not match this approach), even if they have the ability to make short-term tradeoffs without thinking through any abstract arguments.

- \* If your values and the values of a would-be employer diverge, then there is a serious incentives problem with phrasing the question in this way. If you work for organization Z they are very likely to capture the value you create, whereas if you donate to your favorite organization (from their perspective) it isn't too likely that you will really reliably donate to them in particular (or even at all). From your perspective this is a bonus (since organization Z might not really be your favorite place to donate, it might just be a place where you fit in well as an employee), but from their perspective it completely messes with the calculation.

## **Waiting and learning**

If I suspect that my current "best guess" altruistic activity is not really the most effective way to improve the world, and moreover if I think that I'm likely to have a better best guess soon, then money has another important advantage over direct work: namely, I can easily redirect donations as I learn more, or save money to donate later [LINK]. This appears to be a consideration in favor of earning to give.

However, we aren't learning more because of the passage of time alone; we are primarily learning because of experience and work that is currently underway. "Money now" can be turned into "money later" and so may be preferable to "direct work now," but this is not a strong argument in favor of "money now" as opposed to "learning now." Indeed, the same argument suggests that you should avoid direct work while in a state of sufficient ignorance.

So as long as we think that learning is important and are considering doing direct work, we should focus on direct work that helps with learning. We should think about donating to vs. directly working in that context, and this consideration does not seem to distinguish between earning and working.



## **Flexibility**

A significant benefit of "earning to give" in general is that it is very easy to pivot from supporting one project to supporting another, and that the skills that are developed in traditional and lucrative careers are much more likely to remain useful than skills developed for a particular project.

For example, suppose that I think the most important project is project X, and part of the reason why I am enthusiastic is that I expect that after we've put 100 man-years into project X we can reevaluate and pivot to either project Y or project Z. If I do work on project X myself, the capital I build will be much less valuable after this pivot. But if I earn just enough to pay for a replacement to work on project X (which we might expect if my replacement and I are exactly interchangeable) I will build skills which I can later apply to earning money for project Y or project Z.

If the best way to learn at the moment were to pursue a project which we might well later abandon, then I think this would be a huge consideration, and might on its own be enough to carry the argument in favor of earning to give. At the moment I do not think this is the case---I think the most important work to be done for learning is likely to be very flexible and general (for example, research of the sort GiveWell or CEA is doing), and that indeed much of the learning will take the form of capital developed while pursuing those projects. So at the moment I think consideration of career capital broadly favors working directly: see the next point.

## **Skills and learning**

A major benefit of pursuing any career is developing skills, knowledge, and connections. Working directly on X is likely to produce useful skills for working on X in the future, or knowledge about whether X is the best thing to do which can't be trivially transferred to someone not working on X, or connections that will help with similar projects in the future or will help evaluate whether X is a good thing to do.

The value of any of these resources is higher for someone who cares about X than for someone who does not, and so this consideration will tend to favor doing direct work. It is not clearly distinct from the "big consideration" #2 above.

## **Overhead**

When we talk about a worker who is  $\frac{2}{3}$  as productive, it is important to emphasize that

this does not just mean a worker who does  $\frac{2}{3}$  as much in each hour of work. There are other resources that can limit an organization or movement's ability to take on new people, such as the overhead of managing people, training new people, or coordinating amongst people. In many contexts these kinds of overhead are very important, and so a worker who does  $\frac{2}{3}$  as much is much less than  $\frac{2}{3}$  as valuable---indeed, you might not be willing to pay for them at all!

This consideration makes quality more important when evaluating productivity for direct work, but (prima facie) also makes salary more sensitive to quality, so it's not clear on balance in which direction it should point.

I would guess at the moment that it cuts gently in favor of doing direct work if you are unusually productive and earning if you are not particularly productive, because larger organizations are better equipped to handle the overhead of training / coordinating / managing and so are more willing to take on these costs.

That isn't a very robust view though, and certainly can be easily overwritten in particular cases.

### **Performance elasticity of salary**

If salary in a career is not very elastic to performance, the picture given above does not change. However, *if* low-productivity workers will tend to pursue jobs where salary is inelastic (we might expect this), then you should not expect such workers to have comparative advantage in X.

### **Flexibility as comparative advantage**

If there are very large imbalances in the trained labor force (for example if there are way too few trained programmers) than merely being early on in a career may give someone a strong comparative advantage, by allowing them to cheaply retrain and pursue a neglected discipline. In this case, it might be better for young altruists to pursue neglected areas while providing funding for specialists to work on the problems they care about (because those specialists have already invested significant time in developing those skills).

This requires (1) a sufficiently imbalanced workforce that this consideration is relevant, which should be embodied by large wage disparities, and (2) candidate hires who have built up career capital that is particularly applicable to X, but which is disfavored by the

current labor market.

Anecdotally, condition (1) may obtain today, though it is a bit tough to document this effect given unobservable variation in employee quality. (And I am not very confident in this judgment.) Condition (2) seems less likely; the kind of work that I perceive as important (e.g. careful, critical research) does not seem to be substantially disfavored by the existing labor market. But software engineering and finance stand out enough as high-earning areas that it is quite plausible that there are important types of work that are in much less demand.

This consideration looks different from the others we have discussed, because it doesn't materialize as an obvious comparative advantage in some lucrative career---indeed, someone might have basically no experience or particular affinity as a software engineer, yet still be justified in training, if demand for software engineering was high enough. (Though to repeat, the situation is a bit more subtle than it at first appears: this only works if you can hire someone who *would* have to retrain to earn as much as a software engineer, but *would* be able to make use of their current skills to work directly on X.)

### **Inflexibility as comparative disadvantage**

Relatedly, if I want to do something over the coming decade that involves 30 years of experience, I'm much better off paying someone else to do it (even if I as a donor can only cover a small fraction of their salary).

### **One vs. many**

If a task can more cheaply be done by a larger number of less-skilled workers, or a smaller number of more-skilled employees, than I might not have comparative advantage in X even if I am better at X than anyone else who has comparable earning potential---instead, either less-skilled or more-skilled workers should do X.

## **Conclusion**

Overall, I now feel quite comfortable suggesting that an altruist ought to pursue a high earning career and pay someone else to do X, rather than doing X directly, if and only if they have comparative advantage in the high-earning career. Of course, a third alternative(e.g. doing some other kind of direct work, or funding something else entirely)

might be better than either of these.

This conclusion seems to accord surprisingly well with common sense, especially relative to my previous view on this question. This conforms to a general trend in which my views on many questions have become increasingly normal under reflection (though in some cases this must be in part attributed to the pressure not to seem crazy).

In the current environment, I think this conclusion fairly strongly supports doing direct work, except for people who are particularly suited for a lucrative career, because of the advantages from working towards an objective which you care about directly. Again, this conclusion and the caveat are in somewhat surprising accordance with conventional wisdom.

Naturally, if many people who are passionate about a particular project are unable to work on that project due to funding shortfalls, then this choice will not be available to them. Some people enthusiastic about the project will need to earn money to fund work on that project; all we have to say on that subject is that those people best suited to earning should be the ones who do the earning. We also haven't said anything about working directly on X, for which there is funding, versus earning to support Y. That comparison depends on the relative importance of X and Y.

## **Appendix - Case Study**

For the would-be CEA employee currently in banking, they have two questions:

1. Supporting CEA directly vs. supporting some other cause. This document doesn't speak to this problem (it couldn't, since that depends on facts the document doesn't mention).
2. Supporting CEA financially vs. by working for CEA. This is what the document talks about.

For part (2), the document assumes that what CEA would do with more money is hire people. If you would buy supplies, or pay a consultant, or whatever, the analysis is largely the same (but if CEA needs only a little a bit of some specialist's work, and it would be expensive to acquire those skills only to work for CEA, it becomes very likely that a specialist will have comparative advantage).

The decision is significantly complicated by events taking place over time. At first I'll ignore this, and just consider a permanent career choice (e.g. assuming that career capital is not transferrable between the two domains). I'll get back to it later.

For simplicity, I'm going to call the person making the choice X.

Here is the analysis as I was indicating in that document, which I might revise in practice. I would be happy to talk through more specific details, and will be doing something similar with Ben.

-----  
A:

The most important step in this case is probably understanding the quality of CEA's marginal hire. You can start with CEA's current marginal hire, but things are changing reasonably rapidly and I think we mostly care about the future, so this may not be a great approximation. This is hard to figure, but is probably a good thing to have general predictions about in any case. It's also probably better to focus not on CEA's marginal hire, but about the marginal person doing similar work on EA projects. Nevertheless, for simplicity I'm going to keep saying "CEA" as a stand-in for the broader project.

My best guess is that in the long term the marginal hire will be someone who is on board with effective altruism but not extremely committed to the project, that they will be paid a broadly comparable amount to what they would earn in a conventional career, and that they will be reasonably well suited to the work (roughly as well-suited to the work as the typical white-collar employee). To the extent you have a different guess you will have a different conclusion.

B:

Then we get to the characteristics of X. Again I'll make some representative guesses which hopefully will indicate how I'd approach the question. If X can earn a very good income in banking but would not be a standout employee at CEA, this is a strong indication that relative to marginal CEA employees X's comparative advantage is in banking. In the more typical case, one possible procedure would be to estimate how well X would do work at CEA, and how much income other people doing work that well would typically earn. (This might be very straightforward, if you have reference people who you can compare to the would-be hire and whose current salaries you know. My impression is that CEA doesn't yet have this information because right now applicants need to be really enthusiastic. This makes it a bit harder to evaluate what things will look like in a future where that has changed, but you can still use indirect methods.) This can then be compared to a realistic estimate for X's projected earnings in banking.

There are other approaches to try to estimate relative well-suitedness, and it seems best to use a few. On the other side, we could look at people who are pursuing banking roughly as well as X, and try to estimate how well the best of them would do in work with CEA. Another alternative is to directly compare the characteristics of marginal hires and X to see which is better suited to the direct work, though again this is more complicated if you expect the nature of marginal hires to change a lot in the future.

C:

The judgment above needs to be modulated by the banker's enthusiasm about CEA as a project and how much advantage this would give them in doing the work. If the sense is that being a dedicated EA confers a very large advantage, and if the analysis above came out somewhat indifferent, then this is a strong argument in favor of pursuing EA work. In general it will be harder to do the comparison (in principle both figures can be rolled into multipliers on productivity, and that is probably how I would try to think about it). Ideally this would be factored in at step (B), by influencing your judgments for how well X would do work with CEA compared to the marginal hire.

D:

If X has comparative advantage in banking, then they expect to be able to get more of their equivalents to do EA work by taking a job in banking, and should do that. If X has comparative advantage in EA work, then paying someone who is as good as them is likely to be more expensive than they can afford on a banker's salary, and so they shouldn't.

----

Caveats and clarifications:

0. Absolute quality doesn't much matter, except insofar as it is predictably more important either in banking or with CEA. If someone earns more income by virtue of being more effective, *prima facie* a similarly effective replacement is correspondingly more expensive to hire (and conversely).

1. A big uncertainty I see is how enthusiastic the marginal hire is about CEA as a project. I was assuming they were unlikely to be very enthusiastic, because the enthusiastic people are relatively likely to be solid hires at *some* job that is doing good directly. If you think this isn't the case, it significantly increases the bar for doing direct work. In this case not everyone who is enthusiastic about CEA as a project can do direct work, and some of them need to earn. The question then becomes more like: *amongst* the population of future people who are enthusiastic about CEA as a project, is X particularly well-suited to banking? The cutoff is determined roughly by the ratio between the number of people from this class pursuing direct work vs. supporting those pursuing direct work. E.g. if that ratio is 2 : 1, then the top third of income-adjusted people, ordered by comparative advantage in earning, should earn.

2. Adjusting income for quality of life. Typically in order to maintain a fixed quality of life a banker needs to spend more money on themselves, and this will be reflected in differences in the salary someone has to be paid to be a banker or work with CEA. When we talk about "income" in a job we mean to talk about the gap between income and spending to reach this fixed quality of life.

3. Various frictions make it harder to donate. These apply to everyone who is giving to CEA, so don't seem to make much of a difference in the decision.

4. Things changing over time. Here I was mostly considering the decision "work for CEA forever, or bank forever" (where "forever" means just until you get enough information about the career path to bail on it, i.e. without a plan for bailing after some amount of time) while the real decision may be more subtle. In particular, if CEA's demands for resources change significantly over time then it may be better to change paths as those demands change.

The most obvious way things are changing is that right now CEA's marginal hire may be more enthusiastic about effective altruism than they are likely to be in the future. This makes it harder to have comparative advantage in working with CEA *right now* than it will be *in a decade*, and so increases the probability that you should earn now and then work with CEA later.

The effect of this consideration on the analysis depends on how costly it is to swap paths. If it is impossible then you should run the analysis above. If it is free then you can run the analysis above using current marginal hires to decide what to do now, and using future marginal hires to decide what to do in the future. In the more realistic case you do something in between the two. Probably the direction of change will be "earning ----> direct work," so the relevant case is the one where direct work looks good

enough in the future to be worth doing now if it was a permanent commitment, but isn't good enough to be doing now.

The great majority of one's earning potential (and I suppose also of one's potential impact in other endeavors) seems to come later in a career. So I think it is reasonable to view the decision now as a decision about what kind of capital is most likely to be useful in the future, and then to make decisions based on that general goal. This is offset by the possibility that stuff-now is more important than stuff-later for CEA, at a rate that is comparable with personal development over a career. So in principle you'd want to compare those two rates of growth and see which one wins, in order to understand whether most of your effect was coming from stuff in the near term or stuff in the long term.

5. Building career capital. This is another orthogonal consideration. Above I was describing a choice between careers and assuming that each career was better at building capital relevant to that career. I think that is a very plausible assumption; the skills, info, connections, status, etc. gained from working on a particular project seem most transferrable to other similar projects. But of course if one path is better than another for general career capital, then this might change the above analysis.

6. Changing attitudes; being around CEA might have a positive effect on a person, moreso than it would on someone who would be hired to do the job instead. Right now I'm calling that a wash.

7. People irrationally don't like banking. There might be some extra oomph from pursuing banking, if many people take a moral stand against it and hence accept lower wages than they otherwise would. It should be pretty clear how to factor this into the above analysis (if someone could do a job in banking but won't for reasons that don't apply to X, then we can look at the gap between what their salary would be and what it is, to evaluate how much of a disadvantage that preference is for their earning).