Note from Holden: I wrote up this doc to try to express some rough thoughts to people I work with a lot. It's not really intended for broader consumption, but I'm OK to share it along with the disclaimer that it's not carefully expressed. Because it's not carefully expressed, I think it would be easy to misinterpret; accordingly, I consider most of its usefulness to come from the potential that it might affect your inside-view models of things, and I think it will often be a mistake to give the things I say here much weight for "trust/deference" reasons.

I see my main role in life, over the last several years at least, as trying to figure out who should work on what. In the course of doing this, I've absorbed and/or developed a sort of worldview - a set of pretty deep and constant assumptions - that inform how I think about questions like "How can I cause event X to happen in the world?" and "Should I bet on person P to do a good job in this particular case?"

I suspect it's not tractable to share the assumptions in a precise enough way that I can really convey them, or convince people of them in a vacuum. But it's possible to gesture at them, and I've decided this is worth trying to do. This is partly because putting a bunch of gestures and hypotheses out there might cause readers to accept versions of these views as they make their own informal observations that seem to support them. It's also partly because **I want to gesture at a "kind of worldview" and a "topic to think about" that feels like it comes from a different intellectual ontology from what EAs are used to.** The message of this document is less "Believe the things I say" and more "Think about the things I think about; have opinions on them. If you do this, you'll likely find that your opinions tend to be pretty inarticulable and unamenable to 'valid argument' style communication. You'll be engaging in a kind of intellectual exercise that isn't much talked about and probably isn't much done in EA, but plausibly is at least as important and valuable as the kinds of intellectual exercises that EAs are more used to." (To be clear, I don't think this will be totally new to anyone who's been a manager, and I don't expect it to blow anyone's mind, but I think it could be a useful nudge in that direction.)

The DRI-centric worldview

Here's an attempt to gesture at a worldview that I think is pretty common in Silicon Valley, especially among people with a lot of management experience. I don't endorse it universally and unqualifiedly, but I think it's a really good lens to have in your toolkit, and I think I'd generally bet on it when it's directly conflicting with something else.

"DRI" is a Silicon Valley acronym - "directly responsible individual" - referring to the idea that if you want something done well, you should designate a DRI for it. This person isn't *necessarily* the person who is "in charge of it" from a power perspective (though Daniela and I think it usually should be, and in fact often dislike the term "DRI" in corporate contexts for this reason - we think DRIs should be owners/managers), but it's the person who is "directly responsible" in the sense that the thing going well or poorly is on them. I chose this term even though I often dislike its use, just as an evocative shorthand.

The DRI-centric worldview in a nutshell is:

- Serious impact in domain or task X is nearly always the product of a person who is obsessed with X and has spent a lot of time on X (relative to other people, and ~always over ~1000h even for very green-field X).
    - The rough mechanism is that any impact on the world requires understanding and dealing with a large # of little things. You need someone who is sharp and adaptive, but most importantly puts in the time and focus to deal with all of these things.
- So if you are trying to have impact on X, you'd best either become that person or recruit/develop/manage them; other things are unlikely to work. If you're noticing that X isn't going as well as you hoped, your first and possibly last question should be whether the right person (or people) is working on it.
- The game of figuring out who is a fit for what is a contender for "best thing to be good at." This is a great thing to think about and build knowledge of.

Some corollaries, elaborations, predictions and implications

- I think the most concrete and consistent prediction of this worldview is that you aren't going to see a lot of exciting impact from casual interventions, by anyone, except insofar as such casual interventions consist in directing meaningful resources to other people.
    - I think a lot of people expect that small/casual/side/temporary projects can change the world. This is especially true for people with sequence thinking habits. I think this kind of thing pans out less than anyone expects who isn't using the DRI-centric worldview.

    - Note that the DRI-centric worldview predicts that even superstars like Elon Musk will fail to have much impact or insight in areas they are devoting little time to. (Except, again, via transferring their resources. But not via their ideas or efforts.)
    - The DRI-centric worldview provides some support for the idea that one should "use common sense as a prior" (in certain respects) and "avoid running too many experiments at once" - whatever you're trying to change about the world, you're probably going to run into a lot of little obstacles, so you probably want to use a lot of normal best practices (a low-overhead way of avoiding unknown unknown obstacles) and focus your bet on the thing you are trying to change.
    - The DRI-centric worldview is generally pessimistic about the impact of advice, even from luminaries, at least when that advice is being directed at people who are already competent and deep in something.
- The DRI-centric worldview advocates for a bit of an obsessive, often beyond-what-seems-reasonable dedication to the idea that everything that matters should have an unambiguous point person, usually someone who has a heavy concentration of both power and responsibility w/r/t whatever X they are working on. (This person is then held directly accountable for how things go.)
    - Interacting with tech founder types, I've often found it almost surreal the way they automatically go "Well X is in charge of Y and I'm not going to second guess them - we are going to do what they want." It's hard to say why it feels so surreal without actual examples, but like ... a lot of times it seems really obvious that the person is crazy and doing a certain kind of thing wrong ... and yet over time I feel like their attitude gets vindicated.
        - In this interview with Reid Hoffman, Reid essentially says (paraphrasing) "The board can be at red light, yellow light or green light on the CEO; if it's at red light that means you're firing them; if it's at yellow gotta get to red or green really fast; if it's at green do what they say." I think this is a striking way of thinking about boards, which are really designed and intended to distribute decision making power across multiple people and serve as a "check" on the CEO. But, I think it's a pretty reasonable view and I can believe it's backed up by experience.
    - Another thing I've been struck by is how consistent it is that "complex reporting relationships" and "confusing ownership situations" (e.g., it's confusing to describe who is responsible for what and/or who reports to whom) are ~never sustainably good, no matter how logical they seem at the time.
- The DRI-centric worldview is really obsessed with commitment. Anytime someone seems like a sure thing to obsess over X for the next 10y of their life, the DRI-centric worldview is tempted to bet on them to succeed. Anytime someone seems like they've got one foot out, the DRI-centric worldview is like "This is going to suck." (An exception would be when a person is on the way out of something they've already mastered - like a CEO leaving a company - and is helping transition.) I think the DRI-centric worldview is less surprised than other worldviews by how little mark various brilliant, competent, scattered people have made on the world.

- The DRI-centric worldview thinks people should be pigeonholed, should have a thing, should be able to sum up their entire life and place in the world in an easy to understand sentence or two. (If someone is really into what they do they should've figured out how to elevator pitch it effectively.) The DRI-centric worldview might be a significant source of memes that are generally skeptical of anyone whose work is hard to describe or sounds very "general."
- The DRI-centric worldview really doesn't like the idea of trying to "have an idea" that one hands off to another person to execute. It likes the hybrid visionary/executor.
- The DRI-centric worldview has an interesting attitude toward "experts": it will always bet on the expert over the inside-view argument, but also sometimes has a weird definition of "expert." If you can establish that you're truly more dedicated to X and have spent 10x more cycles thinking about it than people who seem more superficially like "experts in things related to X" - even if those people are far more prestigious - you can often convince the DRI-centric worldview to totally ignore the "experts" and bet on you.
- Large organizations are conglomerates of large #'s of people, so it isn't necessarily the case that one person can change them (no matter how good they are and how much power they have). The people and the habits are more important in many cases than the incentives, the resources, etc.

In general, I feel (in a way that I couldn't substantiate well without more work) that I've had a lot of surprises in my life that have updated me toward the above points. There are lots of times when other worldviews, and basic internal logic, is excited about some project because the idea seems so good. But when the key ingredients outlined above are lacking, it generally ends up bad.

Related: http://www.paulgraham.com/genius.html

Some contrasting worldviews

Here are some ideas I've seen in various contexts that I consider to be straightforwardly contradicted by the DRI-centric worldview. (Some of my goal with this section is to dispel an impression that the DRI-centric worldview is vacuous or trivial.)

- People trying to make early career decisions based on how much impact they will have in the next 1-2 years. I find it bizarre when Open Phil hires are intent on helping the org in the first weeks. The DRI-centric worldview says you need big-time dedication and investment to an area before you can expect to be having big impact.
- People who overestimate the returns to casting a wide net (getting 1000 resume submissions; putting out an RFP; announcing a prize) relative to getting referrals & asking around for who's good.
    - A lot of times (not always), asking around gets you 80% of the people who are "already super into X" or "a super good fit for X." These people tend to know each other and be findable. The other 20% count for something, to be sure ... but you're going to get especially little value from "randos with a fresh perspective"

(unless your goal is to test how your ideas go over with randos, or your goal is to pull more people into obsessing over your thing in the future).
    - The DRI-centric worldview is a significant source of my skepticism about prize philanthropy.
        - What matters is whom you get to work on your project.
        - What you don't want is people who randomly jump into an area to take a potshot at your prize. What you do want is people are obsessed with the relevant thing. A lot of times, your prize is likely to attract a lot of the former, while doing little to increase the supply of the latter. When there is a credible argument that a prize will get participation from people who are already obsessed with something, or will cause true sustained shifts such that people will get obsessed, I am more positive on the prize.
    - Similar arguments apply to prediction markets. If you can get existing experts to play, or if you can create the right kind of conditions (high and consistent liquidity; some kind of prestige and/or recognizable career track or "industry") that you really expect people to make big investments in learning to play, while fully expecting those investments to be worthwhile, then I expect success. If not, I don't, and think a lot of prediction market bulls are expecting "magic" that I don't expect (or at least, on timelines that I don't think are realistic).
- Similarly, people who think that putting stuff out on the public web for "input from the world" is going to lead to great things. In general I think comment sections are fairly close to worthless in aggregate across all comments posted everywhere on the Internet in the last 10y, and this is way different from what most people's (including, say, Holden2007's) naive expectations would be.
- I see the main narratives/debates in the intellectual politics world to be between a couple of competing worldviews:
    - The fundamental unit of the world is incentives. Get the incentives right and you will get the results you want. Government has crappy incentives so it will do things badly. (Right/libertarian)
    - The fundamental unit of the world is intentions. Put a disinterested government agency in charge and you will get more prosocial results; leave it to the market and you won't. (Left/progressive)
    - The DRI-centric worldview kinda sneers at this debate and notes that a lot of times, switching something from public to private and back has ~no effect. There are plenty of government projects that work perfectly well and can compete effectively with private versions; there are plenty of government projects that are as evil as any corporation could be. Yes, incentives and intentions matter, because they shape who ends up working on something and how their actions are constrained; but there are also all kinds of random things that affect whether good people end up working on something, and these often go totally orthogonal to the markets/government and incentives/intentions axes. DARPA is bold and innovative like the stereotype of a great company; Microsoft is sclerotic and bureaucratic like the stereotype of a government; prosecutors are venal and

corrupt and narrow-minded like corporations (until progressive prosecutors win elections and suddenly they aren't); I'd say Google is high-minded and benevolent to rival a European government (and surpass the US).

- Related to above - economists have a rival worldview that I think has a lot to recommend it, but also misses a lot. I think there is a lot you can't explain except by talking about the people, cultures and habits in a company/country. I think when reading economists trying to explain stuff, this often reads as kind of an entertaining blindspot. And yes of course it can be modeled - ~any decent worldview can stretch to accommodate ~any view - but it is "unnatural" to put this stuff into economics frameworks, and I think it shows.
    - Random thought, but I think the DRI-centric framework has kind of a unique take on the efficient markets hypothesis. If you ask, "Is my stock trading idea good enough to beat the market?" the answer is "Are you in the top echelon of sharp, adaptive people who spend the most time thinking about how to beat markets? If so, yes; if not, no." The DRI-centric worldview predicts that the efficient markets hypothesis is effectively "true from the perspective of a day trader, false from the perspective of Renaissance Capital." I can't say this is a frame economists never formalize or take, but I can say I haven't seen it, and that this implies it's not the most natural fit with their frameworks.
- I think rationalists and EAs often expect to get really far by thinking things through using a particular methodology, with an emphasis on logic. I think in practice the main positive things they have produced look like "Thinking about something ~no one else thinks about, obsessively and dedicatedly." Meanwhile, rationalist/EA attempts to improve on things that lots of other people do with passion and dedication look pretty lackluster.
    - One pattern I think has held pretty well in my experience is that you should beware of people who seem to have a lot of epiphanies. Major changes in one's "operating algorithm" (which is a more important thing about a person than their "beliefs," according to the DRI-centric worldview) should usually be driven by lots of little lessons they are accumulating; if a blog post represents a huge update (especially on some general/vague topic about how to live, as opposed to a more specific point about what problem to work on), I think this implies that there is something wrong with their ability to accumulate and retain little things day to day, and/or their propensity to weigh the little things they've picked up vs. words they're reading.
- The DRI-centric worldview dislikes "halo" reasoning - expecting X to go well because amazing person Y is on it. If it doesn't have their full attention and fit their wheelhouse, don't expect much.
    - I think Elon Musk, Jeff Bezos, and others of that ilk have a "halo" theory of themselves, leading to things like Blue Origin and The Boring Company. Peter Thiel said something like "never bet against Elon Musk" - I would be happy to bet against Elon Musk on one of his lark projects, though not on something he's giving 50%+ of his time to.
- Relatedly, the DRI-centric worldview is kinda bearish generally on "advice," which I think is points for the DRI-centric worldview. If you want advice, you should want advice on X

and get advice from someone who knows the hell out of X, and scope things in a way that you're really going to be hearing about what the person knows about. If you ask someone about your situation, they don't really know about your situation, and no matter who they are their usefulness is going to be highly questionable. (Ideally they're self-aware enough to warn you against putting too much credence on their best-guess course of action.)

My thoughts on how to win at the DRI-centric worldview's game

The DRI-centric worldview thinks the best thing to be (obsessively, dedicatedly, persistently) good at is matching people to roles. But *how* do you do this well? I put much less credence in my views on this than in the DRI-centric worldview overall (the DRI-centric worldview overall is, I think, kinda shared by a lot of credible seeming people, whereas my views are just my views), but here are some notes.

Early career

The DRI-centric worldview thinks the basic equation to have impact on X is ability (general) + adaptiveness (general, though often people are more adaptive where they feel more ownership, which is correlated with what they're excited about) + time (this is where persistence and interest in X come in). (From above: "You need someone who is sharp and adaptive, but most importantly puts in the time and focus to deal with all of these things.")

Early-career people haven't figured out their "thing" yet, which makes them pretty hard to evaluate. You can't look at their track record, and with some exceptions you often can't look at what X they're into at the moment and assume it's persistent such that they can be judged by their aptitude for X. But you can:

- Assess raw ability/impressiveness
- Assess a slightly broader notion of "general capabilities." I am definitely more positive on early-career people when they have normal, conformist "achievements" (eg grades, often proxied by whether they went to a good school) under their belt; this shows the ability to generally "notice that I should be achieving some X, and figure out whatever I need to in order to achieve it," which is relevant to executive function and adaptability. I don't consider this a dealbreaker by any means, especially for just-out-of-college people (here the only signal on them is often what school they went to, which just reflects how they performed in high school, which is a super messed up environment that is super early in life). But I think someone with a "conventionally unimpressive resume" probably

has some kind of limitation/blindspot, at least for the moment, and this is more true as someone is later in their career.[1]

- Assess the "basic capability to pick something to get into and approach it with persistence and/or ownership and/or adaptiveness." It's a good thing if someone has previously "owned" some project that they took from start to finish and was some sort of reasonable contribution to the world on its own terms; this tends to require persistence/ownership/adaptiveness.
    - A red flag for me is if someone was or is super into some X, but has a giant blind spot around it, or some aspect of it they really should've learned about and haven't. For some reason, the guy who runs [redacted] comes to mind here (though maybe he'll make this section look dumb in the future). He demonstrates an incredible amount of a certain kind of passion for [why the thing he's promoting is theoretically good]; but I've like never once seen him argue that [the thing he's promoting] has led to a specific good outcome, or could lead to a specific good outcome. This feels to me like a weird blindspot that means this guy isn't really into his X in the right way. There's a lot of stuff like this. I think people tend to be too quick to judge early-career folks based on dumb stuff they say on random topics; all early-career people are dumb about a lot of stuff. But I think it's much more legit to judge a person for being off on something they really "owned."
    - I think people who either are, or quickly become, good at the "meta procedures" for avoiding really big bad mistakes are very promising; people who are in a domain long enough that they really ought to have started being good at these "meta procedures" with respect to that domain, but haven't, worry me. (I have a short doc elsewhere on "meta procedures" and "meta mistakes" but I'm not sharing it at the moment. But some examples in footnote.[2])
- Assess the general kind of X the person might be able to tackle with the needed approach. I think excitement/passion is quite relevant here, but far from the only thing, and it also needs to be interpreted with taste:
    - If a person is excited about AI in the abstract, this doesn't mean they're a fit for anything re: AI.

---

[1] However, it's important to know that I don't consider weird jobs to be "conventionally unimpressive." I'm using "elite California" standards, and I think those are quite friendly to working for weird nonprofits, weird startups, etc. rather than Bain. They do expect that someone has a plausible story for why they wanted to work at weird organization X, and some kind of stories about what they contributed and accomplished there.

[2] Illustrative examples of "meta procedures":

- If X is important, there should be a person who has unambiguous ownership of X and is the right person for this (which includes things like having been assessed at doing X, being enthusiastic and attentive about X, and having a good network of other people to get advice from on X).
- Before taking an important action, you should consult the right people
- Before causing someone to plan around you, you should model out your ability to deliver
- Generally always try to commit to less, and get more feedback/new evidence on how something's going before you take the next step.

    - It's better to look at things the person has actually put a lot of time into, and ask what "kind of thing" they are - whether they tend to involve reading people, analyzing messy empirical things, analyzing tight conceptual arguments, etc. ("what kind of work").
    - I generally try to ask a person what they've been most proud of and enjoyed most in the past; visualize what kind of day to day work it was; try to get them to visualize stuff they can work on today, such that they're thinking about what it would be like day-to-day; and see what they seem excited about. I think this has signal.

In general, you should assume that an early-career person is many pivots away from finding their optimal X (although they might rack up some wins along the way if they're particularly strong).

Note: I tend to look at very early-career people as people who need to try a bunch of stuff and eventually develop into people who can have big impact. I think this is right for the majority of people. But I think the standard YC lens expects impact right away, and as such puts a higher premium on being intense about some specific thing.

Late career

When thinking about late career people, I think a decent approximation is: "This person just does the same thing over and over." I think people tend to overestimate the generality of late-career people, across the board, and if you cartoonishly imagine them as some kind of algorithm played on repeat, you're probably updating in the right direction (even though this isn't totally right).

If you can find a late-career person whose X is like 80-100% correlated with the thing you want them doing, and who has achieved impressive things with X in the past, you should bet on them over anyone. But as that correlation drops, the prognosis worsens dramatically.

I think there are some exceptions along the lines of "People tend to evolve from doers to managers/investors over time." So for example, OP's scientific program officers spent most of their lives as scientists, but can now apply the patterns they've seen to judging high-level people and directions; I think this is a common pattern for academics/intellectuals.

Finally, I do think there are people who can learn new things late career, but you should think of this as an exception, and something you can specifically test for pretty explicitly. Like, I feel like a lot of late-career people just aren't interested in learning new things at all anymore, and if you try to get them to do anything that doesn't fit with what they've done in the past (even if it's pretty easy) all kinds of stuff will break, and then you know not work with them on anything that involves learning new kinds of things.

For fun I wrote up a paragraph interpreting myself as a late-career person who just does the same shit over and over, but I'm leaving it out for now.

Wheelhouses

Somewhere in between early-career (trying lots of stuff, leveling up general skills, learning high-level stuff about oneself) and late-career (doing the same thing over and over) is the "search for the wheelhouse."

A wheelhouse could be a contained project and a huge burst of energy (there are some domains in which I think a lot of the impact come from someone catching fire for a year, and then kinda burning out - I think Aaron Swartz did this a cpl times, and I think you see this with novels and music a lot), or a discipline that someone sticks with for a decade-plus before they have any impact, because that's what it takes (I think some of academia looks this way).

I think the basic ingredients of someone being in their wheelhouse are:

- They have the general capability to do the day to day work at a high level
- They feel "ownership" over what they're doing - there is some way in which they really understand and buy into the goal they're trying to achieve
- They are paying attention to any little thing relevant to hitting their goal, and accumulating zillions of little lessons (adaptiveness).
- They have the persistence and/or burst of energy needed to do a lot of work in some sense (a "burst" wheelhouse probably means a crazy amt of work in a short period of time; a "persistence" wheelhouse means a normal amt of work over a pretty mind-numbing amount of time; there are in-betweens).

Hopefully this general equation/mindset makes it clear why I think excitement/passion are quite important to pay attention to, though far from the only such thing.

Finding people's wheelhouses is one of the things I most enjoy and find most fascinating. I don't have any kind of interesting formula for it; I just think a lot about someone, what they've done that seemed "most in their wheelhouse" and most impressive, what they've done that seemed worst and that we should make sure they don't have to do again, and what worthwhile tasks remind me of what they've done well ... and I bounce lots of stuff off them looking for the "snap" where their excitement surges, their "ownership" surges, and sometimes (not always) their time spent surges. That sounds very simple and anti-intellectual, but to me looking for wheelhouses feels very fascinating and intellectual, just not in a super explicit/analytical way.

One exercise I've found useful for myself in deciding what I should work on is asking myself questions along the lines of the following (though I'm not sure I've really used this to help other people find their wheelhouse):

- If I could be world-class at anything I put my mind to, what would I put my mind to? What's something I could imagine working really hard on, while having very few serious

moments of wondering whether I could've done something else really well instead? ("If I can do this well, it doesn't really bug me what else I've left on the table")
- What's something I could imagine working really hard to be good at, and then if I failed, feeling like "Dang, I failed, I suck" instead of "Eh, this is kind of a silly game anyway"?

The "value/opportunity" evaluation framework

I'm pretty skeptical of judging people negatively for doing or saying dumb things. The DRI-centric worldview tends to think that everyone has a lot of capacity for being dumb and silly, and that the way people become sharp and reasonable is by having a domain of competence that they get to know well.

I'll judge someone negatively if they do/say something dumb that does *real damage*, and to a lesser degree, if they do/say something dumb *in a domain they really ought to have mastered by now.* Doing real damage relates to the bit above about "meta mistakes" - something about a person not knowing what domain they should be sticking to and/or not mastering the domain they are trying to be in.

But otherwise, I'm way less interested in a "demerits" or "mistakes" framework for evaluation than in a "What have you accomplished, and what have you had the opportunity to accomplish?" framework. People who haven't had the opportunity to demonstrate what we're hoping for are often worth giving a shot, even if they haven't produced much to date, as long as there's a plausible story about how their strengths and weaknesses are well suited to some X they've never tried (and ideally never had the opportunity to try) before. People who have had a lot of opportunity to have impact on X, and haven't, probably won't, even if you can reinterpret all their mistakes as "things a reasonable person might have done." This is because marinating in X for a long enough time ought to cause someone to internalize lots of little lessons and thus become effective; it's overly generous/deferential to excuse someone's lack of impact because you could see yourself making any given decision the way they did.

(I think this evaluation framework has advantages over a "what mistakes have I seen this person make?" framework both in terms of predictiveness and in terms of incentive compatibility. I like people to know both that (a) they don't need to be worried about "slipping" or saying something silly in front of me, as I'll generally shrug that off; (b) if they have a year to work on X and don't produce some awesome X, that's a big negative update, almost no matter what their story is.)

Evaluating people as a manager

Most of the above is about pretty arm's-length evaluation: looking at people's resume, past, etc. This is often what we're stuck with (in recruiting, usually in philanthropy, etc.) But if you manage someone or otherwise work hyper-closely with them you can run some more "controlled experiments."

My usual formula here is:

- Start by asking them to do something that you have a maximally clear plan for evaluating the quality of. It should be something such that if they do a good job, you'll know it and be impressed; if they don't, you'll know it and downgrade your estimate of how useful they are in that domain.
    - For sufficiently senior folks, this can be very open-ended stuff. You just need to be clear on what you're expecting ("Send grant ideas that I'm excited about"), and whether it's a fair expectation given what they've done before and what context they already have.
    - I usually believe in optimizing first assignments almost exclusively for this criterion, and giving ~zero weight to direct value of the assignments, except insofar as assignments with some minimal amount of direct value tend to be more representative/predictive than assignments with none.
- If they don't do well, give them something more scoped and better defined in the same domain, or comparable in some other domain. Again, have a strong plan for evaluating it.
- If they do well, and repeat this once or twice, you have now identified "something this person can do," and can probably ask them to do very similar/seemingly identical things again with a stronger assumption that the product will be good. However, at some point you will probably want to stretch/challenge them further, and at that point it will become important again for having a clear plan to evaluate.
- I generally try to give people a mix of tasks that test different kinds of skills. I also generally try to mix well-scoped/defined work, where I expect good performance and will lower my assessment of the person if they do poorly, with more open-ended work where strong performance is a "bonus" and a failure to do anything doesn't update me that much.
- Maintain a "baseball card," informally in your head or formally in a Gdoc, that keeps track of:
    - What this person's best and worst moments have been - particular things they did better than expected, or worse. (I like to focus on things they did, in the sense of having power and responsibility for the things going well, not e.g. conversations that hint at particular qualities they have.)
    - Hypotheses about this person's powers, anti-powers and competencies.[3]
    - Hypotheses about what this person's wheelhouse might eventually be.

It's important to have a good model of what kinds of things "get better with very directed practice/feedback" vs. "get better gradually largely via osmosis" vs. "aren't going to get better."

[3] Powers =~ "This person is really good at this kind of thing. Every time they work on it, good things happen." Anti-powers =~ "This person is not good at this kind of thing. Every time they work on it, good things don't happen. They struggle." Competencies =~ "This person can do X fine when needed, but it's not their favorite thing or something they really go above and beyond at."

My model of this stuff is largely intuitive and I don't have a ton of concrete stuff to say about it at the moment. But some broad thoughts:

- If someone's background is in X and their job seems to have required being good at X, you can expect them to be good at X, and if they're not, that's a really bad sign for their general promise/competence, unless you have a very specific hypothesis about how things will go differently this time.
- However, if you're asking someone to do stuff for your org that doesn't rhyme heavily with things they've done before, you should be pretty patient. There are a lot of things about a given org that take time and practice to get better at, and a good chunk of these things happens via gradual osmosis.
- There is some kind of "ownership leap" I've seen people make that consists of something like:
    - Marinating in a culture or domain for a while.
    - Being in charge of something going well, at some point, with no one there to back them up who seems knowledgeable/attentive enough for them to defer to.
    - Suddenly becoming much better at navigating open-ended situations, as a result of really being in charge of something and being qualified to be in charge of it and feeling that they are qualified to be in charge of it (and realizing that nobody else is going to catch them if they fall).
    - I'm not really that great at reliably "manufacturing" this leap, but I do generally expect that someone needs time to marinate in some domain and have their hand held for a while before they can make it, and that they also need the training wheels taken off and to be left totally in charge of something with very little backup before they can make it.

Creating and assigning roles

"Creating and assigning roles" is a thing that I consider very interesting and to have a lot of room for variance (w/r/t someone's ability to create and assign roles effectively). It's relevant to both philanthropy and management (you can think of stuff to fund people for, and think of stuff to assign people to).

I think a normal way of thinking about roles is in project/task terms: "I need someone to do X, who can do that?" I think an alternative way of thinking about roles that is often better (esp for philanthropy, as opposed to management) is something like: "There should be someone who thinks about X all day and does whatever is needed to make X go well, perhaps including hiring their own team. Who could that be?" or "I need something to change w/r/t X; there is already a solid specialist in X; how can I give them the incentive/interest/understanding to be motivated to push in the direction I want?"

I have kind of a dance going on in my head between the things I wish people would do and the people I know exist (both specific people and "types"). The story of Open Phil's role development is very bidirectional: I try to figure out what I could see someone obsessing about

for their whole life, I think about what Open Phil needs, and I try to find some intersection. And I keep revisiting my picture of possible roles over time and try to re-slice it such that the roles are more likely to someday get filled.

You're rarely going to get someone to do exactly what your inside view says they should do. You should feel excited about people who seem good and are doing something vaguely in the area of what you want done, or thought about. An awesome person doing something that kinda sorta rhymes with what you're excited about, and could kinda imaginably end somewhere like what you're excited about eventually, is arguably more exciting and worth betting on than an OK person doing exactly what you're excited about.