

The “EA Forum Peer-Reviewed Work” (EAFo-PR)... a proposal in context

- Authors: David Reinstein (and Jaime Sevilla ... to add content)
- Linked/embedded in the [“Unjournal planning gitbook”](#)

Background on the EA forum and “rigorous research”

The [EA Forum](#) is a discussion forum that acts as a hub of communication for people involved with Effective Altruism. As a forum, it has a particularly high level of intellectual rigor, openness, and an emphasis on reasoning transparency (see the [norms](#) here).

Some of the posts on the EA forum are particularly academically and intellectually rigorous. (DR: This is perhaps particularly true in areas involving moral psychology, artificial intelligence risk and animal sentience.) Some research institutions and think tanks (Rethink Priorities, perhaps others) use the EA forum as one of the main outlets for their work. This includes very long posts which often involve research projects taking as long as six months to complete. Researchers from other EA-aligned and -adjacent institutions (GPI, FHI perhaps) sometimes post links to their work on the forum—“linkposts”—sometimes with short excerpts or discussions..

Forum posts get a number of types of feedback, including upvotes, comments and eligibility for certain prizes (*most important posts of the decade*, etc.). However, on the Forum, all participants essentially count equally, and an upload from someone with little familiarity with the topic will count as much as an upvote from someone who has considerable expertise in an area.

The EA community—and major funders and decision makers—are particularly interested in understanding the credibility of the arguments and evidence presented in Forum posts. This can be challenging. Only a small set of those posting original research and rigorous arguments are part of the “academic publication system” or seek structured validation from experts in other areas. There are limited ways to get this validation outside of the cumbersome and differently-aimed traditional journal system (for reasons related to those discussed in the *Unjournal* writings).

The Proposal: the EA Forum Peer-Reviewed Work

Jaime Sevilla (and others?) have proposed and discussed ways in which the EA Forum could facilitate better feedback and “credibility assessment,” particularly in areas requiring detailed evaluation, high level technical and academic training and expertise, and particular domain and subject matter knowledge.

Note: this is something like the XXX in the LessWrong forum.

Essentially, the proposal involves a separate section connected to or within the EA Forum that focuses on this type of work. We would use standard (?) academic and automated (see Elicit) tools for finding and recruiting expertise. We pay (e.g., \$500 per review?) field and subject-matter experts to write specific “referee-report style”

assessments of the work. We may also need some funds to pay “editors” to manage this process.

We are considering (and have some differences of opinion over) whether

- these reviews should be anonymous (perhaps with the reviewer’s “credential” indicated) or the reviewer’s names should be posted.
- the reviews should be quantitative assessments, comparing the work to other published works, perhaps making predictions about its “publication success” or other outcomes.

There are pros and cons to both of the above. Posting one’s name along with one’s review may lead to better credibility and increase the reviewers’ incentives to put in care and effort. On the other hand, people may be reluctant to be negative in contexts where authors know their identities, and this may make reviews less informative. Reviewers may also be more comfortable being anonymous for other reasons. For example, they may not want to be specifically associated with effective altruism or a ‘less-than-standard-academic’ context. however, to the extent this is the case, reviewer’s “coming out in public” will increase the credibility of EA and this outlet, and should thus be encouraged.

Quantitative assessments and metrics (if done right) have certain advantages. so they can be compared to one another, and can be used to consider and model the relevance and truth value of certain propositions. David Reinstein believes that quantitative and comparable assessments are necessary if this is to be a good example that could form a bridge to academic work and bringing academics into the fold. (Some discussion below, and in ‘Unjournal proposal’. He also thinks this is important forgetting reviewers to be definitive in their evaluations, in a way that will be useful for assessing the credibility. Jaime Sevilla has some reservations about these. [JS – I’ll let you expand on this

Connection to the ‘Unjournal’

This proposal comes from a different direction as the “unjournal” project. The EAFo-PR involves “upgrading” less formal and less academic work that would typically be posted on the EA forum. In contrast, Unjournal proposal aims directly at academics and career researchers, encouraging them to submit their work for evaluation before, during, or after the standard ‘traditional journal’ mode. That proposal is partly about improving research and academia, and to some extent about making connections between academics and ideas and topics of interest to effective altruists. EAFo-PR aims to make the connections in the opposite direction.

My (DR) impression is that the best and most involved work on the EA forum is of a very high reasoning standard and sometimes includes and incorporates substantial expertise. In some ways, including communication and reasoning transparency, I think it often exceeds traditional academic work. The forum’s openness and the nature of feedback given there often allows a lot more input and more of certain types of scrutiny than the traditional academic process. In some fields in areas, only a few people will actually ever read a published academic paper carefully (sometimes not even all of the long list of authors).

However, in general terms posts on the forum do not involve as long a period of effort and careful honing. Even the (rare?) top end of forum posts probably only take a few months to prepare, while academics can work for years on a project and open print at least in economics the period between a first working paper and ‘publication’ is typically something around six years. Furthermore, academics do you have years of specific training and follow techniques “standing on the shoulders of giants” in ways we may not typically see in forum posts.

I think we can also expect a certain amount of skepticism about the rigor of forum posts both in policy circles and particularly among academics. Academics may be hesitant to have their work associated with “a mere post on a social forum” either from their own concerns, or from second-order beliefs about how it will look to others.

Thus, I don't think the EA Fo-PR is likely to be the “same outlet” as what we have in mind for the Unjournal. I think it may be difficult to get full-time academics themselves to submit work that is closely tied to the EA Forum... partly because of the name and branding.

However, I do think the EA Fo-PR will be a useful example, proof of concept, proving ground, and learning experience for the Unjournal. It will also be a good way of forming ties and building infrastructure between effective altruism people and research slash academia people, that will be a win in general, and for an EA-aligned ‘unjournal’.

OTHER NOTES:

Jaime Sevilla (FTX ?) has

- Paying reviewers is probably the best way to get this done. Some people might volunteer in EA but it won't scale.

'- Overlay' (like Lesswrong) ... Posts can be submitted to the 'EA journal' ... brought this up with Aaron Gertler Identify the audience and how to serve them.

- Reach out to EA forum authors, offer to review their papers, connect them to reviewers

Alt: 'hey academics, we offer to review your papers, and will review those of others'

Write up a post in the EA forum ... a CtA ... this comes from the 'upgraded forum post' side, and could help us understand how it works.

Thinks that ranking/rating is overhead that is not worth it. ... thinks this can be oversold in the prediction context. Getting reviewers on board is going to be harder. Also worried about Goodhart's law.

- Does this paper do X

Issue with 'stars' ... very anti-rating ... one star means "blah blah blah"

Anonymity? we can discuss. Slightly against blind reviews. 'Name attached' gives an additional incentive. Wants 'professional reviewer' to be a thing.