# Certificates of Impact

Paul Christiano Follow
Nov 29, 2014 · 9 min read

## 0. Introduction

In this post I describe a simple institution for altruistic funding, based on the creation and exchange of "certificates of impact."

Typically an effectiveness-minded altruist would try to do as much good as possible. Instead, users of the certificates system try to collect certificates for as much good as possible.

Whenever anyone does anything, they can declare themselves to own an associated certificate of impact. Users of the system treat owning a certificate for X as equivalent to doing X themselves. In the case where certificates never change hands, this reduces precisely to the status quo.

The primary difference is that certificates can also be bought, sold, or bartered; an altruist can acquire certificates through any combination of doing good themselves, and purchasing certificates from others.

For example, a project to develop a malaria vaccine might be financed by investors. If the project succeeds the developers can declare themselves owners of an associated certificate, and distribute it to investors in the same way they might distribute profits. These investors could then resell this certificate to a philanthropist who is interested in funding malaria prevention and who honors certificates. The philanthropist would recognize the certificate as valuable to the extent that they believed that the original project was causally responsible for the development of the vaccine, and their willingness to pay would be the same as their willingness to develop the vaccine themselves (evaluated with the benefit of hindsight).

Note that judging the value of a certificate is just as subjective as judging the value of the underlying activity, and is done separately by any philanthropist considering acquiring that certificate.

## 1. Some existing challenges

1. Evaluating philanthropic opportunities is difficult, and most are not very good. Predictions are expensive and inaccurate, and often infeasible for small donors. The problem is worst when evaluating novel interventions. Prizes can address some of these difficulties, but have their own set of problems.

2. One person might have the ability to do a project, another the desire to fund it, and a third the knowledge to evaluate it. Coordination is hard, and existing incentives misaligned, both for disseminating good information and executing good projects.

3. Thinking about crowding out and causal responsibility is hard: if I don't do something, will someone else? Does it matter which of Oxfam's activities I fund? How is causal responsibility divided between donors and employees? Between the philanthropist who buys malaria nets and the government which distributes them?

4. Prices are an elegant and flexible system for communication and coordination, but they are often unavailable for altruists.

5. Reasoning about leverage, and interacting with funders with different priorities, is tough. It's easy to end up neglecting a large possible upside or engaged in zero-sum conflict.

# 2. Certificates of impact

Whenever anyone does something good in the world, they can "mint" an associated certificate of impact and declare themselves to be the owners. This can be as simple as making a public statement, analogous to acknowledging a funder. At the bottom of this post I could write "This post backs certificate [RationalAltruist-22], which is now owned by Paul Christiano;" I could do the same in a press release describing a medical intervention, an academic paper reporting the results of an experiment, or a piece of open source software. Independently, I might post an offer to sell these certificates, or a philanthropist might contact me and express interest in purchasing them.

By deciding which certificates they are willing to pay for, altruists using the system define the rules of the game. The protocol suggests considering a certificate valuable only if

1. It was issued by the group or individual that performed the associated activity.

2. It is the unique certificate associated with that activity.

Users should be willing to pay the same amount for a valid certificate of X that they would be willing to pay to cause X to happen (and to keep the resulting certificate). That is, I should be indifferent between spending $100 to do a good deed and paying someone else $100 to do it, provided that in either case I get the entire certificate for that good deed.

Naturally this willingness to pay is subjective, depending on both values and views—just like conventional grantmaking. Users maximize the total value of all of the certificates they hold, using whatever combination they prefer of doing good deeds themselves or acquiring others' certificates.

Certificates can be transferred just like any other IOU or title, and have the same practical issues with bookkeeping and double-spending. We can resolve these however we like (word of mouth, a bulletin board, a trusted central party, the bitcoin blockchain…). Establishing properties [1] and [2] probably involves trusting the issuer—just like conventional grantmaking.

Certificates can be subdivided arbitrarily. When a group accomplishes a project together, they must decide how to divide the resulting certificates in the same way that a profitable enterprise would decide how to divide the profits.

Owning a certificate for X would probably not carry the same prestige as doing X yourself. Funders using the system purchase certificates as a way to do good, not necessarily as a way to earn credit (though helping them get credit may be useful for getting them on board).

For now, I'm most interested in the question: would philanthropic activity in an area be improved, if many funders used this system or a similar scheme?

# 3. Commentary

## Some simple examples

**No transfers.** In the simplest possible example, certificates never change hands. In this case, the protocol reduces to "Do the activities that you think have the largest positive impact on the world," i.e. the status quo.

**Grants.** A grantmaker might pay a charitable organization to perform some charitable work, in return for the resulting certificates. This also reduces to the status quo, at least for the grantmaker. A grantmaker might also leave some of the certificates to the charities, in the same way that a startup retains equity as an incentive.

**Prizes.** An unfunded do-gooder might keep the resulting certificates to themselves. Later, a philanthropist who appreciates the value of their work could buy the certificate off of them. The end result is very similar to a prize or bounty. As with a prize or bounty, that philanthropist could state their intentions in advance, choosing their own balance between flexibility and predictability.

## So what?

Reproducing the status quo isn't so exciting. Do certificates improve the situation, or even change it? I think so, in ways that should be interesting to particularly effectiveness-minded funders. For example:

1. Allocating certificates requires explicit and transparent allocation of causal responsibility, both within teams and between teams and donors. In addition to the obvious cultural effects, this aligns individual incentives with altruistic goals, reducing incentives to mislead donors, volunteers, and employees. Overall I would be excited about these effects, though there are *significant* costs.

2. Exchanging certificates for money leads to a more consistent conversion between good done and compensation, especially if funders use a mix of prizes and grants. It can prevent good deeds from falling through the cracks, ameliorate some winner-take-all PR dynamics, and eliminate some zero-sum conflict. It also makes it easier to move between different funding modes.

3. Purchasing certificates helps funders with different values coordinate; if I see a good deal I have an incentive to take it, without worrying about whether it might be an even better deal for someone else.

4. The ability to resell certificates makes a purchase less of a commitment of philanthropic capital, and less of a strategic decision; instead it represents a direct vote of confidence in the work being funded.

There are also risks inherent in bringing altruistic incentives closer to profit incentives (substituting extrinsic for intrinsic motivations, introducing new scope for zero-sum conflict…). How you feel about the tradeoff depends on how you feel about the relative merits of the two approaches, and may vary across domains.

## Some more examples

In addition to these simple examples, certificates of impact can facilitate more exotic interactions.

**Financing.** If I'm optimistic about your project I might give you a loan which I expect you to pay back by selling a certificate. Or I could buy the certificates off you, as if I were a grantmaker, with the intention of reselling them.

**Research universities.** A research group can employ researchers in exchange for ownership of some of the certificates of impact they produce. These certificates can then be sold to funding agencies or philanthropists. This shields those funders from having to make predictions about research output, which they are often not equipped to do, and provides good incentives to both researchers and research groups. (This is a special case of financing and causal attribution, but it's an application close to my heart.)

**Interpolating between prizes and grants.** By implementing both prizes and grants as trades, it is easier to interpolate between the two, providing some funding to get a project going in exchange for part of its output (presumably at an expected discount) and then purchasing more of its output after the fact.

**Moral trade.** Many cases of moral trade can be implemented by turning good deeds into certificates and bartering or trading in a marketplace. For others, such as abstaining from eating meat, this would be problematic (though it's not clear that the use of certificates is to blame).

**Targeted donations.** If I want to fund only part of an organization's activities, I can purchase only those certificates of impact which I find valuable.

**Research on effectiveness.** If I think a certificate is too cheap, I can do research to evaluate its effectiveness; depending on the result I can buy the certificate, publish my research, and resell the certificate.

**Trades across time.** If my discount rate is lower than interest rates, I can sell certificates of impact today, invest the proceeds, and purchase certificates next year. (The price of each fixed certificate should rise at roughly market rates; but in addition to that, a certificate for a life saved in year 1 may have a different value than a certificate for a life saved in year 2, reflecting the discount rates of philanthropists.)

**Speculation.** If I think at a change will be considered good in retrospect, I can purchase appropriate certificates now and aim to resell them later at a profit.

**Multiple bottom lines.** If a project has several objectives, it can use certificates to monetize all of them and then make local profit-maximizing decisions. This may lead to more sane decision-making, and much more transparency, than the status quo.

(I'm sparing you a much longer list.)

## Some theory

Suppose that everyone who valued X were using the certificate protocol. (For concreteness, suppose it's everyone interested in cosmology research.)

Then at equilibrium, the price of certificates of impact on X is equal to the marginal cost of achieving an impact on X. Any deviation is an arbitrage opportunity: if you can do X more cheaply, then you can sell the resulting certificates for a profit; if you are doing X more expensively, then you could save money by buying certificates instead.

Moreover, the status quo corresponds to one way of producing certificates, with total value equal to the total good being done. The actual outcome should be a Pareto improvement, and in particular should increase the total value of available certificates. As a result, it also increases the total amount of good being done.

I wouldn't lean on this argument alone. But I do think it suggests that things might tend to work themselves out, at least if the system were widely adopted. For me, considering concrete cases bolsters that intuition.

The infra-marginal units of X, those that were cheaper to produce than the marginal unit, pose a theoretical problem. These units of X

get produced and purchased (good news!), but the resulting certificates are most likely to be sold for the marginal price of X (bad news!). That means there is a transfer of wealth, from the people who care about X (and who use the certificates system) to those who have infra-marginal opportunities to do X.

This problem is worst for the early adopters of the certificate system, which is particularly unfortunate. And it's even worse when the opportunities to do X are so cheap that they were going to get done anyway. For example, suppose some high energy physicists would have to go out of their way to avoid doing some cosmology. Under the certificates system they would get paid for this research by cosmology funders who purchase the resulting certificates. But the cosmology funders might not be so happy about this, since the counterfactual impact of this funding is zilch.

Whether this is a feature or bug depends on how many really cheap opportunities to do X were failing to get done. Maybe the high energy physicists wouldn't even bother publishing their cosmology in a format that cosmologists would understand, because the high energy community wouldn't give them any respect for it. In this case there might be huge social benefits from paying them to do so. My own guess is that we leave a lot of low hanging fruit hanging, and that picking it is more important than worrying about the transfers.

# -1. Note

I have omitted any discussion of how certificates might be implemented, especially discussion of how you might get from here to there. This is a tough problem, but I think that a single advocate could make meaningful headway if it seemed worthwhile. I'm interested in the prior question, is it a "there" worth thinking about?

. . .

*Originally published at rationalaltruist.com on November 15, 2014.*