A Note on *The Good Woman of Szechuan* Paradox: A Fundamental Problem for Effective Altruism

Walter Barta

Winter 2022 (DRAFT)

TLDR: The fundamental problem at the heart of Effective Altruism is the tradeoff between effectiveness and altruism, exemplified in the Bertolt Brecht play "The Good Woman of Szechuan."

Ever since seeing it performed as an undergraduate, I have been thinking about Bertolt Brecht's play "The Good Woman of Szechuan," and the problem that its story poses for the would-be good Samaritan. Here I meditate on my interpretation of the conflict of the play, and how it relates to the Effective Altruism movement. Chiefly, I consider how many arguments against effective altruism have been put forward, and many are serious, but only one argument against effective altruism is fundamental, the conflict between effectiveness and altruism illustrated in "The Good Woman of Szechuan".

Effectiveness/Altruism Incompatibility

We can make critiques of effectiveness itself (e.g., cluelessness, etc.), or we can make critiques of altruism itself (e.g., rational selfishness, etc.), but these critiques will always be generalized to effectiveness and altruism respectively, not to effective altruism particularly.

A critique specific to effective altruism will pit effectiveness against altruism. And if effectiveness can be altruism are shown to be incompatible, this will pose a fundamental dilemma for effective altruism particularly. The fundamental problem would have to look like this:

> **General Effectiveness/Altruism Incompatibility**: One can either be altruistic or effective, but not both.

But does the Effectiveness/Altruism Incompatibility arise in all cases? Surely not, as it seems that many examples of effective altruism compatibility can be pointed to.

So, the incompatibility would be conditional, arising only in cases in which there are 1) potential increases in effectiveness and altruism, and 2) tradeoffs between effectiveness and altruism. Essentially, incompatibility arises anytime one is attempting to optimize for both effectiveness and altruism, creating a non-trivial multivariate optimization problem, finding oneself on a Pareto Frontier, where every improvement in some dependent variable will cause a decrease in some other dependent variable (Miettinen, 1999). Thus, the more limited version of the dilemma would look like this:

> **Conditional Effectiveness/Altruism Incompatibility**: One can either be altruistic or effective, but not both, under certain limited conditions, specifically, if one finds oneself on a Pareto Frontier in a multivariable optimization problem.

So, perhaps certain normative systems can avoid the Effectiveness/Altruism Incompatibility under those conditions in which it arises? Possibly, but possibly not.

Unfortunately, both conditions (1 and 2) that together are sufficient for the Conditional Effectiveness/Altruism Incompatibility are fundamental positions of Effective Altruism (Deere, 2016).

First, Effective Altruism advocates the position that "helping more is better than helping less" (Deere, 2016). Any kind of normative system that requires the optimization of certain variables (e.g., classical utilitarianism, etc.) will run the risk of the Effectiveness/Altruism Incompatibility. But, worse still, any system that advances the increase in some normative variable (e.g., welfare, etc.) or the decrease in some normative variable (e.g., illfare, etc.) without bound, like welfare-increasing and suffering-reducing systems, will run-up against optimums at some point, and thus risk the Effectiveness/Altruism Incompatibility.

Second, Effective Altruism advocates the position that "our resources are limited" (Deere, 2016). Because we live in a competitive ecology, with many different organisms and institutions vying for constrained resources, we perhaps always exist on some Pareto Frontier or other, making any normative system that advocates for increases also an normative system advocating for some other decreases.

Thus, the Conditional Effectiveness/Altruism Incompatibility seems to be baked into the premises of Effective Altruism, at least all versions of Effective Altruism that accept 1) normative improvements and 2) resource constraints.

The Good Woman of Szechuan Example

To give an example of the Effectiveness/Altruism Incompatibility at work, we can look to the Bertolt Brecht play "The Good Woman of Szechuan," which has a quite compelling portrayal of the tragedy of philanthropy. The titular Good Woman of Szechuan notices the appalling conditions of the world and wants to do something about it, so:

1. The Good Woman decides to try to improve the world.
2. But, in order to improve the world, she needs money.
3. So, in order to get money, she runs a business.
4. But, in order to succeed at business, she must be competitive.
5. And, in order to be competitive, she cut costs.
6. And, improving the world is one of the costs that must be cut.

So, the plot of the play ends up being a sequence of actions done with good intentions that become self-defeating. Her means undermine her ends, and the Good Woman ends up merely perpetuating the appalling conditions that she was attempting to improve—albeit with the benefit of some tragic-ironic self-awareness. The Good Woman finds herself needing to optimize profit in order to optimize charity. However, optimizing profit is in tradeoff with optimizing charity (on the Pareto Frontier of Szechuan's marketplace), and the basic structure ends up being: in order increase altruism, one

must increase effectiveness; but, in order to increase effectiveness, one must decrease altruism. Thus, this example of Conditional Effectiveness/Altruism Incompatibility we can colloquially call the "Good Woman of Szechuan Paradox."

On one interpretation, the Good Woman's vicious cycle is plausibly the vicious cycle that the infamous Sam Bankman-Fried, Effective Altruist turned Crypto-Scammer, found himself in (Parloff, 2021). Before the collapse of his fortune, Bankman-Fried may have had sincere intentions to improve the world in the spirit of Effective Altruism, and so he followed the path of the Good Woman, running a business in order to make a profit in order to give to charity, but in the end his means undermined his ends, and he ended up making the world a worse place. To put Bankman-Fried's problem in terms of Effective Altruism, there may be a tradeoff between being effective (trying to make a ton of money for charitable giving) and being altruistic (actually giving away the money to worthy causes). Of course, other issues, like irrational risk-taking and Machiavellianism may have been at play in the Bankman-Fried case, so we cannot make these judgements conclusively, but we can at least imagine Szechuan-like cases in the modern world.

Brecht's own ideological commitments are Marxist, and he thus sees his play as a parable about the need for revolution to replace the capitalist system, which would resolve the incompatibility between effectiveness and altruism by relieving the constraints of competitive market conditions, ushering in a cooperative system in which no tradeoffs arise. If Brecht is correct, then the Effective Altruist movement must inevitably embrace some form of revolutionary systemic change, something that the movement is reluctant to embrace, often framing itself as reformist and incrementalist in disposition. But, like other revolutionary arguments, we may find Brecht's unconvincing, if we believe that a condition of limited resources will persist even beyond capitalist systems. In as much, the Good Woman's problem, not as Marxists frame it but as we may reframe it, may be a deeper problem than capitalism or socialism, perhaps as deep as normative theory itself. If resource constraints and competitive pressures afflict all states, then any Utopian state must be afflicted too. Viable Effective Altruist Utopias would be ones in which effectiveness and altruism trivially coincide without optimization tradeoffs, but whether or not such a coinciding of effectiveness and altruism exists in the space of possible Utopian states seems to be an open question. Instead, we are left with having to find some acceptable tradeoff points; being the most effective we can be while still being altruistic, or being the most altruistic we can be while still being effective, suboptimal for both effectiveness and altruism, but perhaps still conditionally possible.

Either way, if there is any problem at the heart of Effective Altruism, we should bet on Conditional Effectiveness/Altruism Incompatibility, exemplified by the Good Woman of Szechuan Paradox. Fortunately, the conditions of Effective/Altruist incompatibility do not obtain under all circumstances, making some Effective Altruist activities possible. Unfortunately, the conditions do obtain sometimes (perhaps oftentimes), leaving us on a frontier of tradeoffs, wondering if we are doing the right thing.

Brecht, Bertolt (1942). *The Good Woman of Szechuan*.
http://www.socialiststories.com/en/writers/Brecht-Bertolt/The-Good-Person-of-Szechuan-Bertolt-Brecht.pdf

Sam Deer (2016). “Four Ideas you Already Agree With (That Mean You’re Probably Already on Board with Effective Altruism.” Cited by *The EA Handbook* from *Giving What You Can*. https://www.givingwhatwecan.org/blog/four-things-you-already-agree-with-effective-altruism

Kaisa Miettinen (1999).Nonlinear Multiobjective Optimization. (Kluwer/Springer). https://bayanbox.ir/view/2515773690068372592/Kaisa-Miettinen-Nonlinear-Multiobjective-Optimization.pdf

Roger Parloff (August 12, 2021*).* *"Portrait of a 29-year-old billionaire: Can Sam Bankman-Fried make his risky crypto business work?"*. *Yahoo!Finance*.