Philanthropy and Impact Evaluation

Paper prepared for the International Philanthropy Research Conference
22 and 23 September, Turin, IT

Gian Paolo Barbetta, Centro di Ricerche sulla Cooperazione e sul Non-profit**,** Università Cattolica del Sacro Cuore – Milano

Contents

Introduction

This paper deals with the issue of "philanthropy" and "impact evaluation", the object of several pieces of literature in recent years (Center for Effective Philanthropy, 2016; Coffman et al., 2013; Coffman and Beer, 2016; Ebrahim and Kasturi Rangan, 2014; Greenwald, 2013; Kinarsky, 2018; Kinarsky and Christie, 2022; Williamson, Leat and Scaife, 2017). Nonetheless, before moving to the core of our discussion, we need to specify what we mean by each of these terms. Both expressions could take on different meanings in different contexts.

First, philanthropy could generally be defined as the attitude of an individual to promote the well-being of others, especially by the donation of money. This attitude (together with its opposite, egoism) is as old as humanity. Nonetheless, this paper will not focus on individuals, nor will it consider the reasons that motivate them to donate, which is the topic of specific literature in the fields of psychology, sociology, and economics. Instead, I will only consider organised philanthropic activities undertaken by modern grant-making foundations (GMF, from now on). Therefore, the focus of my attention will be on today's private organizations, most of which were established in the second half of the twentieth century, whose endowment generates the income they use to make grants supporting private and public organisations that act for the common good.

While the main organizational characteristics of these institutions are similar in several countries, some differences have developed over time and in the various states. For example, while in the past most GMFs were established in perpetuity, nowadays many of them have a finite organizational horizon and, besides their income, they also spend their endowment out. Moreover, in some states, GMFs enjoy very favourable fiscal rules while in others they are subject to full taxation. These differences notwithstanding, the main structural characteristics of these organizations are the same across western societies. The reason for this focus is that the number of philanthropic grant-making foundations has greatly increased over recent years, and they now play a relevant (and sometimes controversial) social and economic role in most western states, particularly (but not exclusively) within the wide field of welfare service provision.

Second, when talking about impact, I will refer to a change in a "state of the world" (measured by a specific target variable) that occurs *because* a specific action was deliberately taken by someone (a GMF, in our case) and that would not have occurred if the action had not been taken. Therefore, I will talk about the *causal* link between an action and a change in the state of the world. In this respect, I refer to the model developed by Rubin (1974), the so-called counterfactual (or "potential outcomes") framework (Holland, 1986). The counterfactual framework provides the only rigorous definition of the impact of an action on a target variable, but its empirical application (and, consequently, the activity called impact evaluation) is far from easy. We could measure whether the world changed (by measuring the different levels of a target variable over time), but we can't measure the *causal link* between a specific action and the change in the target variable, which is an intellectual construct, not an empirically observable object. Said differently, given that we cannot observe (and therefore measure) what would have happened if the action had not been taken (the counterfactual), we cannot *measure* impact.

No reasons to be sad. Over the last fifty years, the new science of "causal inference" (Pearl and Mackenzie, 2018) developed a language and methods that now allow *estimating* (not measuring) impact, at least when actions are taken at a population level (which means involving several persons or units).

Therefore, in the following pages, I will not address the quest for a measure of the impact of Foundation X (not rigorously answerable) but I will try to understand whether foundations should invest in estimating the impact of their actions (programs, initiatives, grants, etc.) on final recipients. My answer is that they should, much more than they do, but the issue has several complexities that must be addressed.

The paper continues as follows. Section 2 deals with the mission of foundations; section 3 relates a foundation's mission with impact estimates; section 4 reflects on how impact could be assessed; and section 5 draws some conclusions.

What are GMFs for?

Until recently, scholars have "characterized material support for evaluation by foundations as modest" (Greenwald, 2013, p. 505). On the contrary, over the last few years, it appears that GMFs paid more attention to this issue, at least in the US, but reliable figures on evaluation expenditures by GMFs are still missing (Kinarsky, 2014). The feeling is that the bulk of the philanthropic sector is still in the process of discussing possible merits and disadvantages of evaluation, as well as the different approaches to this practice, but field experience is limited to a few of the largest GMFs (see, for example, Twersky, Arbreton, and Trivedi (2019) for the Hewlett Foundation and Gates foundation (2022)).

Is this good or bad? Should GMFs pay more attention to the issue of evaluation and, most of all, should they devote more resources to evaluating the impact of their activities on final recipients, following the evidence-based approach that is gaining the consensus of only a few of them[1]? Or should they save the evaluation money to increase the resources devoted to the funding of their core activities?

There is no unique answer to these questions; in fact, answers greatly depend on individual beliefs about what foundations are for, what social role they should play in society, their strategic vision, and their specific activities. In principle, and adopting an oversimplified model of a GMF, one could say that such institutions could play four main roles: *charitable*, *advocative*, *developmental,* and *innovative*.

When playing a *charitable* role, GMFs use their resources to relieve the consequences of a problem (whose causes could be naturally or socially determined) on individuals. In this respect, they could give funds to shelter people in case of a natural disaster or feed individuals that cannot re-enter the labor market in a recession. These activities, which try to reduce the negative impact of a social problem by tackling its most relevant consequences, have always been part of the mission of GMFs and quite often they represent the very reason these institutions are established. In recent years, GMFs have started diversifying their activities and adopting different roles, such as advocacy, developmental and innovative ones.

When playing an *advocative* role, GMFs do not use their resources to directly address a social problem, but rather devote their funds to raising public awareness about the problem itself and its consequences. The hope is that, because of increased social sensibility, other actors (such as public administrations or civil society organizations) will address the problem and possibly solve it. Conversely, GMFs play a *developmental* role when devoting their financial resources to increasing the scope of welfare, cultural, educational, and research systems, to directly address the problems of some neglected classes of individuals. In this respect, the advocative and the developmental role can complement each other; GMFs direct some seed money to address problems and, at the same time, raise public awareness about the problem itself, hoping that other actors will enter the field with resources large enough to tackle the specific problem. When playing this role, GMFs (and society) know how the problem should be met, but society is not devoting enough resources to adopt the necessary solutions.

On a quite different level is the *innovative* role. When acting as social innovators, GMFs know very well that the economic resources they can devote directly (or raise from society) to activities aimed at solving a specific collective problem are generally not enough to come to a complete solution. Moreover, and even more important, they are aware that in many fields we simply do not know what the right activity could be.

[1] See, for example, the strategy declaration of Arnold Ventures at https://www.arnoldventures.org/work.

Therefore, they aim not to directly address and solve social problems, but rather to show how problems could be successfully addressed, experimenting with new answers, and assessing their effectiveness.

Of course, GMFs do not need to focus on just one specific role. Quite often, their activities reflect a mixture of these different roles.

Role and impact evaluation

Estimating impact is not crucial for GMFs playing a charitable, advocacy, or developmental role. The very nature of these roles assumes that GMFs already know what should be done to alleviate the social problems they want to address: Foundations know "what works". Generally, GMFs play these roles because, notwithstanding possible solutions to the problem at hand, society at large has not yet decided to adopt them. GMFs try to change social priorities by increasing the resources devoted to the specific activities that could solve the problem.

On the contrary, estimating the impact of their activities on final recipients is crucial for GMFs which play an innovative role. Innovation is, by definition, a new possible solution to a problem. Innovation without impact assessment is useless because society would not know how effective the solutions devised to solve public problems are. Even more, innovations could fail, and the innovative role played by GMFs could be detrimental to society if new practices are not put to test; in fact, self-styled "good practices" may prove harmful when rigorously investigated.

The good news is that GMFs that decide to play the role of innovation (as well as all GMFs), due to their structural characteristics, enjoy a "comparative advantage"— toward public administrations, market companies, and non-profit organizations — in estimating the impact of their projects on final recipients. Foundations can better face the risk intrinsic in innovative activities because they are independent of owners, markets, and the ballot box and because they rely on their endowment to fund their activities. When innovative interventions do not deliver as promised or prove harmful, politicians risk their re-election, public officials their position, and non-profit and market entities their sources of income. On the contrary, foundations risk nothing: Board and staff members are not elected, and annual income comes from the endowment, not the sales on a market. Moreover, the potentially infinite horizon of their activities allows GMFs to adopt solutions that deliver their results in the long term.

While their very existence is not at risk if innovation fails, GMFs' reputation could be jeopardized by failures, and they could lose the citizens' support. The "reputational risk" (Strachwitz and Alter, 2020) could generate what Anheier and Leat (2019) named the "benign fallibility syndrome", so that "it is in the interests of both grantor and grantees to be economical with admissions of less-than-desired outcomes" (Anheier and Leat, 2019).

This is true if GMFs (and their stakeholders) perceive themselves as problem-solvers. Failures are damnation if the institution aims to solve problems. But they are a blessing if an institution wants to find out what works and what doesn't in solving problems. Failures, made public, are very helpful as they prevent other institutions from making the same mistakes.

How to estimate impact

While GMFs' institutional characteristics make them particularly fit to estimate the impact of projects and interventions on their final recipients, this task is far from simple and requires the adoption of a specific strategy, made up of several steps.

First, GMFs should specify what is the exact problem they want to address and what changes in the world (what "alternative state of the world") they want to pursue. Both the initial and the final state of the world should be measured in an incontestable way through a set (as small as possible) of agreed-upon variables and indicators (the outcome variables). Making this preliminary step is crucial to estimating impact: If you do not know where you are headed, you will never be able to determine whether you reached your destination. Quite often, when funding projects and interventions, GMFs declare fuzzy aims (not precise targets) whose reach is therefore hard to measure.

Second, GMFs should adopt an accurate "protocol of action" to tackle the identified problem and reach the alternative state of the world, as measured by the outcome variables. The effectiveness of an intervention could be established only with an accurate protocol description. When performing an intervention, if several different actions are undertaken without a rule determining which action should be taken for which people at what time, it will be extremely hard to attribute a possible change in the outcome variable.

Third, GMFs should adopt a valid method to determine whether the ongoing change in the state of the world was caused or occurred despite the intervention. To determine the existence of a *causal* link between the intervention and the (possible) change observed in the outcome variables, we need to estimate the counterfactual situation (what would have happened should the intervention not have taken place): The difference between factual and counterfactual would be the measure of the effect of any intervention. Unfortunately, counterfactuals cannot be observed and therefore measured, so they should be estimated. This is the reason impact cannot be measured, but only estimated.

We know that the pure difference observed in the level of the outcome variable "before and after" the deployment of the intervention represents a biased and unreliable estimate of the impact of the intervention itself. This difference could be determined by several other factors occurring while the intervention was deployed. A similar bias could characterize the difference observed in the level of the outcome variable for those who took part in the intervention and those who did not. The two groups could have been dissimilar when the intervention was started so that any difference observed after the intervention ended could simply reflect the original diversity rather than the project's effect. Therefore, GMFs cannot rely on these simple heuristics methods to estimate the effect of their action (even though they often do that).

Over recent years, the new science of "causal inference" has developed several methods to produce unbiased estimates of the effects of an intervention on a target outcome variable, as testified by the list of recent Nobel laureates in economics (Banerjee, Duflo, and Kremer in 2019; Angrist, Card and Imbens in 2021) all of whom worked in the field of impact measurement of policies and interventions.

This new science made clear that no one of the methods developed can estimate counterfactual situations when an intervention only involves one unit (person or institution). Therefore, questions such as "what is the impact of remote learning on Mary's (or Charlie's) school performance?" cannot be rigorously answered, and we better stay away from them. The same happens with the question "what is the impact of Foundation X?" Nobody knows what would have happened if Foundation X had not existed, and a shared and indisputable opinion is hard to reach. Each of us may have a view about this issue, as each of us has strong opinions about the impact of Ronaldo (or Messi) on the last football match, but no one knows what would have happened should this champion not have played the game.

This limitation notwithstanding, when projects involve larger populations, several methods can be used to estimate the average impact of the intervention on its final recipients. One of the most reliable methods is represented by the randomized controlled trial (RCT) (Gerber and Green, 2012; Gertler et al., 2016). With this method, a population of targeted units is randomly assigned to two different groups: the so-called "group of intervention" and the "control group". The law of large numbers will make the two groups very

similar, on average, so that any statistically significant difference between them could be ruled out. When the two groups have been created, the group of intervention will be involved in the project whose efficacy is at test, while the control group (not involved in the intervention) will represent the best possible estimate of the counterfactual situation (what would have happened to the units involved in the intervention, had they not been involved).

This technique (or the other so-called non-experimental techniques of impact evaluation (Angrist and Pischke, 2009) should be systematically adopted by GMFs claiming to support innovation. If not, one could not be sure of the effectiveness of the new interventions aimed at tackling problems. In fact, as it has been recently shown in the field of education (Lortie-Forgues and Inglis, 2019), innovation could fail or even prove harmful.

Some final thoughts

Given the two assumptions whose rationality I tried to argue (1. foundations are fit to take the risk of innovation; 2. innovation needs rigorous evaluation), I believe that GMFs should devote a large chunk of their activities to supporting systematic evaluation of private and public interventions and policies aimed at pursuing the public good. They should produce (and make publicly available) evidence on "what works" in fields crucial for social well-being, such as education, health, and social welfare. Not many other actors in modern societies can play this role. An example, in this direction, is represented by Arnold Ventures, whose mission statement declares "We focus on correcting systemic failures through evidence-based solutions." Playing this role would greatly benefit society, particularly in countries less prone to empirically assess the effects of policies and interventions and more prone to ideological discussion.

Nonetheless, and quite paradoxically, only a few *innovative* GMFs adopt the rigorous methods described above when running their programs or ask their grantees to adopt these methods when managing programs funded by the GMF. This attitude gives rise to what could be called the "paradox of foundations' impact measurement": GMFs sometimes waste their resources questing for measurement of their overall impact (Greenwald, 2013), which cannot be rigorously estimated, while they rarely adopt, and sometimes openly refuse, methods that rigorously allow estimating the impact of specific projects targeting large populations (which can be rigorously evaluated). Why is this happening?

In my opinion, two main reasons explain this trend. First, the intellectual achievements of the new science of causal inference are not sufficiently known by society, as well as by the world of foundations. Academia should do more in this direction. Second, the diffusion of paradigms created in the world of finance has increased the quest for simple and readily available impact indicators. In the world of finance, indicators such as the rate of return, or the return on investment are easy to calculate and represent a meaningful guide to investment strategies. On the contrary, as we tried to discuss earlier, synthetic indicators used to estimate impact are completely meaningless (and outright wrong) if the question of causality is overlooked, as often happens. Nonetheless, less than careful investors (and superficial or unscrupulous operators) may be interested in taking a shortcut.

One further question needs to be addressed. When impact cannot be estimated rigorously (because a counterfactual is hard to define) what should GMFs do? This is a situation that foundations sometimes face, particularly when they play the role of the advocate; in this case, they should assess whether their efforts in raising consciousness about a problem are well made. In these circumstances, the whole set of tools used by qualitative evaluation could be used.

References

Angrist J.D. and J.S. Pischke (2009). *Mostly Harmless Econometrics*, Princeton University Press.

Anheier, H.K. and D. Leat (2019). *Performance measurement in philanthropic foundations*. Routledge.

Center for Effective Philanthropy (2016). *Benchmarking Foundation Evaluation Practices*. https://policycommons.net/artifacts/1847350/benchmarking-foundation-evaluation-practices/2593555/.

Coffman, J., T. Beer, P. Patrizi, and E. Thompson. (2013). Benchmarking Evaluation in Foundations: Do We Know What We Are Doing?. *The Foundation Review*, 5(2): 36-51.

Coffman, J. and T. Beer (2016). How Do You Measure Up? Finding Fit Between Foundations and Their Evaluation Functions. *The Foundation Review*, 8(4):1-18.

Ebrahim, A. and V. Kasturi Rangan (2014). What Impact? A Framework for Measuring the Scale and Scope of Social Performance. *California Management Review*, 56(3): 118-141.

Gates Foundation (2022). *Evaluation Policy*, https://www.gatesfoundation.org/about/policies-and-resources/evaluation-policy

Gerber, A. S. and D. P. Green (2012). *Field experiments: Design, analysis, and interpretation*. W. W. Norton.

Gertler, P.J., S. Martinez, P. Premand, L.B. Rawlings, and C.M.J. Vermeersch (2016). *Impact evaluation in practice*. Inter-American Development Bank and World Bank.

Greenwald H.P. (2013). Increasing the Value of Evaluation to Philanthropic Foundations, *American Journal of Evaluation*, 34(4) 504-518.

Holland, P.W. (1986). Statistics and Causal Inference. *Journal of the American Statistic Association*, 81(396): 945–960.

Kinarsky, A.R. (2018). The evaluation landscape: U.S. Foundation spending on evaluation. In S.B. Nielsen, S. Lemire, and C.A. Christie (Eds.), The Evaluation Marketplace: Exploring the Evaluation Industry. *New Directions for Evaluation*, 160: 81-96.

Kinarsky, A.R. and C.A. Christie, (2022). Analysis of Evaluation Policies in the Philanthropic Sector, *American Journal of Evaluation*, 43(2): 175-192

Lortie-Forgues, H. and M. Inglis (2019). Rigorous large-scale educational RCTs are often uninformative: Should we be concerned? *Educational Researcher*, 48(3), 158–166. https://doi.org/10.3102/0013189X19832850

Pearl, J. and D. Mackenzie (2018). *The Book of Why*. Basic Books.

Rubin, D. (1974). Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies. *Journal of Education Psychology,* 66(5): 688–701.

Strachwitz, R. and Alter R. (2020) Improving trust in trusts: introducing the Philanthropy.Insight tool, *Trusts & Trustees*, 26(6): 483–492.

Twersky F., A. Arbreton, and P. Trivedi (2019). *Evaluation Principles and Practices. A Guide to Evaluation at the William and Flora Hewlett Foundation*.

Williamson, A., D. Leat, and W. Scaife (2017), Narratives of performance measurement in philanthropic foundations, *Voluntary Sector Review*, 8(3): 273-298.