Online Appendix
Presidential Address: Opportunities and Limitations for Social Impact from Innovation in Digital Products

by Susan Athey

In this Online Appendix, I review some of the economic considerations that support the argument that digital technology has substantial potential for social impact and is a strong candidate for public and philanthropic investment, as well as for investment of time and resources by academic institutions. I also identify some potential limitations.

I. The Cost Structure of Digital Interventions

Many types of digital technology are relatively low cost to develop, as the primary investments required involve writing software. The advent of artificial intelligence coding tools and the prevalence of high-quality open source software tools have further reduced these costs. Digital technology often has low or zero incremental costs from serving additional users, and distribution is also typically inexpensive, at least once intended users are aware of the solution and motivated to use it. One caveat is that recently developed Large Language Models do incur non-trivial costs from usage.

In short, digital technology often has scale economies that make it well suited for social impact applications, for example for settings where intended beneficiaries may have low income. It is also well suited for simple forms of government or philanthropic funding, as development of the application is the primary cost, and the complexity of funding physical implementation is avoided.

Although building a particular piece of software may be inexpensive, there are still important impediments to developing beneficial software and having it become widely used. First, foundational research, prototyping, and piloting are needed to operationalize a theory of impact into something that actually works. Second, as I emphasize below, ongoing incremental improvement is often needed to make a product or service work well enough to engender widespread adoption "organically," that is, without subsidies or substantial costly effort on the part of the developer of the software.

Third, the products need to be distributed to the constituents they serve. Some products or services may be most effective when adopted or distributed by organizations, such as schools, governments, or hospitals. Some evidence of efficacy is often a prerequisite for gaining the collaboration of implementing organizations. Furthermore, managing the relationships and the adoption process may be time consuming. Products or services that are distributed directly to consumers or service providers face different but related challenges in gaining adoption. To get awareness, they may need to advertise, something that can be expensive depending on the domain. For example, finance-related impact applications may need to compete for advertising space against for-profit firms. It may also be difficult to reach certain populations through advertising.

II. Generalizable Innovation

Digital technology has been described as a "general purpose technology" based on its broad use, capability for ongoing improvement, and its ability to enable innovation in application domains (Bresnahan, 2010). These characteristics imply that knowledge about how to use or improve digital technology in one domain can potentially be transferred to other domains, amplifying its impact.

For example, a demonstration of an effective approach to solving a problem in one domain may transfer to other domains. Types of innovation that may generalize include user interface design, recommendation system technology, or behavioral and motivational approaches that help recipients of services stick to their plans for engaging with a beneficial service.

In the context of the case studies described in the paper, a generalizable findings is the demonstration of the impact of personalization. In addition, the research articles may serve as a practical "how-to" guide for others in future applications, providing examples how experiments were designed and analyzed. In addition, for several of the case studies, the researchers produced "practitioner guides" that make the tools and ideas accessible to decision-makers who can implement them in other domains.[1]

The projects also make use of a variety of general-purpose software tools, including open source software tools such as those for analysis of heterogeneous treatment effects using data from randomized experiments (causalTree, based on (Athey and Imbens, 2016) and grf (Tibshirani et al., 2024)), as well as tools for estimating optimal targeted treatment assignment policies (PolicyTree, described in Sverdrup et al. (2020)). At a broader level, there are many open-source software tools for which advances generalize to most other domains.[2]

III. Tactical Insights versus Generalizable Scientific Knowledge

Even when innovation involves general purpose technology, it may not be obvious that the innovation creates generalizable scientific knowledge. Although many innovations in digital products may appear to involve primarily tactical insights, I argue here that working on new products or new features of existing projects also provides unique and substantial opportunities to create generalizable knowledge.

First, the innovation process allows researchers to search for insights in areas that may not have been well explored in the past. With an interdisciplinary team that might include economists, engineers, behavioral scientists, domain specialists, and others, combinations of ideas from different disciplines have the potential for new discoveries, es-

[1]Some examples include a report on applications of machine learning to behavioral science policy and field experimentation using case studies on encouraging community college students to renew their financial aid and reducing vehicle booting about New York City drivers (Athey et al., 2021*b*); a practitioner's guide to designing adaptive experiments (Athey et al., 2021*a*), and an application that facilitates experiment planning (Athey and Hadad, 2021); and two reports on collaborations with PayPal using behavioral nudges to increase the quantity and quality of charitable giving (Athey et al., 2022; Athey, Koutout and Nath, 2024)

[2]Some notable tools include the programming languages R and Python, as well as libraries for these programming languages like PyTorch for developing deep learning and machine learning models.

pecially when grounded in a real-world problem and the direct opportunity to build and implement a solution.

Second, finding a solution and showing that it works is a demonstration that *something* in the solution has an impact over the status quo. This often allows a researcher to consider scientific questions that go beyond tactical knowledge about the specific problem in the specific context. An intervention helps overcome a social problem, but why does it work? How does it work? For whom does it work? Does it improve all relevant, measurable outcomes, or does it improve some and harm others, and does this vary by subgroup? If one component of the intervention is changed, does it still work? Essentially, a novel, effective intervention is also a laboratory in which a variety of both established and original research questions can be asked and answered.

Another type of question that arises when considering scientific evidence is whether an intervention's effectiveness is context-specific, or whether evidence generated under tightly controlled conditions will generalize in a more complex real-world scenario. When innovation is embedded in real-world implementation contexts, evidence about effectiveness is less vulnerable to concerns about the applicability, practicality, and generalizability of a research finding for real world problems. In the example of the reading contest described in Section III.A of the paper, the success of a specific implementation of a contest and leaderboard yields tactical insights about how to make a contest that is effective at getting engagement, but it also addresses fundamental questions in behavioral science about whether incentivizing student engagement crowds out future enjoyment of reading, or whether it creates lasting, positive habits.

More broadly, innovation in social impact settings may probe directions that are less of a focus for for-profit firms, filling in a gap in the space of innovation. Measuring lasting impact for a social goal is harder than measuring engagement that generates clicks or advertising revenue. In addition, social impact innovation may focus more on ethics and distributional considerations. For example, the case studies in Sections III.A, III.B, and IV.A of the paper analyze efficiency-equity trade-offs.

IV. Academic Researchers and Interdisciplinary Labs as Contributors to Innovation

Certain types of digital technology can be prototyped at the scale of an academic lab or small non-profit. Examples of labs that have designed, prototyped, and implemented digital technology include the Golub Capital Social Impact Lab at Stanford Graduate School of Business (where the author is the founder and faculty director), as well as the non-profit Research Improving People's Lives (RIPL), the non-profit ideas42, the Carta Lab at Stanford University, the Eberly Center at Carnegie Melon University, and the Behavior Change for Good (BCFG) initiative at the Wharton School of the University of Pennsylvania.

In addition to directly producing usable technology, prototyping digital products often serves multiple socially beneficial purposes, including creating scientific research (in machine learning or experimental design methods, or in social or behavioral science) as well as educating students.[3] The social benefits of digital technology targeted at social impact may be orders of magnitude greater than the potential profits of a hypothetical for-profit firm selling the product or service (where in some case the product or service might not ever have the potential for profits that offset development costs). As a result there are untapped opportunities for impact. In such settings an academic lab may be able to build beneficial products that might not be provided by the private market.

Students planning for their future careers often value practical experience that contributes to their human capital and demonstrates their capability. Such students may benefit from participating in academic-led innovation during their studies. Similar to the benefits of typical research assistantships and industry internships, the experience may make academic training more salient and help students make more informed decisions about future careers. More practical, hands-on experience may also have additional value to future employers. In addition, academic labs focused on social impact have the opportunity to place greater focus on broader long-term social impact, beyond what might be practical in a for-profit firm. Students may also appreciate and learn from participating in interdisciplinary teams. Thus academic-led social impact projects may provide educational opportunities surrounding impact that are harder to find in the private sector.

Some challenges occur in digital technology innovation, both in the for-profit startup setting and in the academic context. One of the most common challenges is a lack of data (and a lack of users of the technology whose usage might create such data), something that is important for measuring impact, guiding incremental innovation, and developing personalization technology. This challenge is also an opportunity for research and development, however; methodological research that focuses on smaller datasets as well as on estimation of counterfactual benefits of policies that have not yet been implemented can help improve the early-stage research and development process. Since many early-stage for-profit companies do not have the capability to develop new methods, let alone create and publish research about them, an academic laboratory can potentially fill in a gap that may be less likely to be filled by the private sector.

For example, the small organization aimed at helping women get jobs in the technology industry in Poland in the case study in Section III.A of the paper identified an unmet need in the market for mentoring, but the founders were unsure how to scale it and did not have experimental evidence on their mentoring program that would attract further funding. The Golub Capital Social Impact Lab was able to help develop a scalable product and evaluate it alongside the original mentoring program, providing both immediate value to the organization and generalizable knowledge on interventions that encourage women into tech jobs. The experimental designs for each program were carefully structured to provide statistical power in a small dataset. Similarly, in the case study described in Section II.A of the paper, Stones2Milestones had a promising product, but

[3]Classes in the Action Learning Program (ALP) at the Graduate School of Business at Stanford University are designed to engage students in real-world problems; the Golub Capital Social Impact Lab contributed to three years (2020-2022) of these courses in which students designed digital products to solve problems for social impact organizations collaborating with the lab. Other initiatives that support building new products and initiatives for social impact include the Rustandy Center for Social Sector Innovation Workshops at the University of Chicago, D-Lab Development Ventures at the Massachussetts Institute of Technology, and Innovation to Impact at Yale University.

lacked capacity and experience for innovation and experimentation with recommendation systems. In this case, the Golub Capital Social Impact Lab contributed to the design of the product as well as the design of experiments, comparing alternative recommendation system methods and selecting one that was appropriate for the size of the dataset. The comparison of alternative approaches made use of methods for evaluating counterfactual policies from the recent econometrics literature (Athey and Wager, 2021; Zhou, Athey and Wager, 2023).

V. Collaborations for Innovation and Distribution through Implementation Partners

More broadly, implementation of ideas after prototype and research has been a challenge, and some have called for researchers who conduct field experiments to take this on as the natural next most important challenge for the field. List (2024) writes "One crucial question for the experimental research agenda in the social sciences relates to the scale-up problem: can this idea work at scale?"

Without actually implementing and testing prototypes, it is difficult to build effective data driven products and to refine them so that they work well. But it may be difficult for creators of a prototype or academic researchers to find an implementation partner without showing that it works, since the expected benefits to the partner may not outweigh the costs. This "chicken-and-egg" problem can be difficult to surmount. Some non-profits have been built to address this problem; examples include the What Works Hub, Innovations for Poverty Action, and the Global Innovation Fund.

One promising approach to mitigating this problem is for academic researchers to develop a partnership with an implementing partner. This may work well if an implementing partner has already established a base of users of a basic product or service, and the researcher has an idea of how to improve it or expand its scalability using digital technology. A collaboration has the potential to surface the knowledge required to design effective solutions, and to co-create implementation strategies that fit well with the objectives, constraints, and costs faced by the implementation partner.

Others have observed the importance of collaborations, particularly when a government is the implementing partner. As Chupein and Glennerster (2018) observe:

> A third pathway to evidence-informed policymaking is to institutionalize the use of evidence in organizations, and particularly governments, that have a comparative advantage in generating resources and who tend to work at large scale. The most obvious path to scaling up evidence-informed programs is through the partner with which the evaluation is conducted. Some of the most successful policy influence occurs when an implementing organization that partners on an evaluation is involved in an RCT's design from the beginning to answer its priority questions.

REFERENCES

Athey, Susan, and Guido Imbens. 2016. "Recursive partitioning for heterogeneous causal effects." *Proceedings of the National Academy of Sciences*, 113(27): 7353–7360.

Athey, Susan, and Stefan Wager. 2021. "Policy Learning with Observational Data." *Econometrica*, 89(1): 133–161.

Athey, Susan, and Vitor Hadad. 2021. "Bandit Experiment Application." Accessible at: https://www.gsb.stanford.edu/faculty-research/labs-initiatives/sil/research/bandit-experiment-application.

Athey, Susan, Kristine Koutout, and Shanjukta Nath. 2024. "Choosing the "Right" Default Donation Amounts for Each Donor to Balance Multiple Fundraising Objectives." Available at: https://www.gsb.stanford.edu/sites/default/files/publication/pdfs/other-2024-nov-choosing-right-default-donation-amounts.pdf.

Athey, Susan, Matias Cersosimo, Matt Jarvis, Dean Karlan, Shannon Coyne Rosado, and Leah Rosenzweig. 2022. "PayPal Giving Experiments." Available at: https://www.gsb.stanford.edu/sites/default/files/publication/pdfs/report-2021-mar-paypal-giving-experiments_2.pdf.

Athey, Susan, Vitor Hadad, Dean Karlan, and Leah Rosenzweig. 2021*a*. "Designing Adaptive Experiments." Available at: https://www.gsb.stanford.edu/sites/default/files/publication/pdfs/academic-publication-desiging-adaptive-experiments-2021-mar.pdf.

Athey, Susan, Vitor Hadad, Niall Keleher, Octavio Medina, Rebecca Nissan, Rachel Rosenberg, Matt Schaelling, Jann Spiess, and Josh Wright. 2021*b*. "Computational applications to behavioral science." Available at: https://gsbdbi.github.io/ml_for_behavioral_science/index.html.

Bresnahan, Timothy. 2010. "General purpose technologies." *Handbook of the Economics of Innovation*, 2: 761–791.

Chupein, Thomas, and Rachel Glennerster. 2018. "Evidence-informed policy from an international perspective." *The ANNALS of the American Academy of Political and Social Science*, 678(1): 62–70.

List, John A. 2024. "Optimally generate policy-based evidence before scaling." *Nature*, 626(7999): 491–499.

Sverdrup, Erik, Ayush Kanodia, Zhengyuan Zhou, Susan Athey, and Stefan Wager. 2020. "Policytree: Policy Learning via Doubly Robust Empirical Welfare Maximization Over Trees." *Journal of Open Source Software*, 5(50): 2232.

Tibshirani, Julie, Susan Athey, Erik Sverdrup, and Stefan Wager. 2024. "grf: Generalized Random Forests." R package.

Zhou, Zhengyuan, Susan Athey, and Stefan Wager. 2023. "Offline multi-action policy learning: Generalization and optimization." *Operations Research*, 71(1): 148–183.