Essays on The Very Invisible College: Global Science and African Participation

by

Caroline Viola Fry

BA Natural Sciences University of Cambridge, 2009
MSci Organic Chemistry University of Cambridge, 2010
SM Massachusetts Institute of Technology, 2018

SUBMITTED TO THE SLOAN SCHOOL OF MANAGEMENT IN PARTIAL FULFILLMENT OF THE REQUIREMENTS OF THE DEGREE OF

DOCTOR OF PHILOSOPHY
at the
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

May 2020

Signature of Author ....

Department of Management
15th April 2020

Certified by ....

Scott Stern
David Sarnoff Professor of Management of Technology
Thesis Supervisor

Accepted by ....

Catherine Tucker
Sloan Distinguished Professor of Management
Professor, Marketing
Faculty Chair, MIT Sloan PhD Program

Page left intentionally blank

Essays on The Very Invisible College: Global Science and African Participation

by

Caroline Viola Fry

Submitted to the MIT Sloan School of Management on April 15th 2020 in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy (Ph.D.) in Management

Abstract

Despite globalization, innovative activities remain concentrated in a handful of high-income countries. Leveraging knowledge and resources in these locations through ties in the global network presents opportunities for emerging economies. This dissertation consists of three essays studying the role of international ties in the development of scientific capacity in sub-Saharan Africa. Each chapter helps to uncover a different feature of the way in which, and the scope by which, international ties impact African science, and ultimately facilitate technological catch-up and economic growth. Chapter 1 is an introductory chapter, and chapters 2-4 are specific research applications. Chapter 2 explores the value of international relationships to African scientists leveraging a unique opportunity afforded to some scientists to develop these relationships: the 2014 Ebola epidemic. Chapter 3 studies the spillover impact of the return home of American trained scientists to African institutions. Chapter 4 explores a macro-association between foreign knowledge stocks and African scientific productivity.

Thesis supervisor: Scott Stern
Title: David Sarnoff Professor of Management of Technology

Thesis supervisor: Pierre Azoulay
Title: International Programs Professor of Management

Thesis supervisor: Ezra Zuckerman Sivan
Title: Deputy Dean and Alvin J. Siteman (1948) Professor of Strategy and Entrepreneurship

Thesis supervisor: Fiona Murray
Title: Associate Dean of Innovation

Acknowledgements

I am grateful to my advisors for their consistent patience and support. Thank you to Scott for supporting me in pursuing my passion, whilst at the same time showing me the rigor needed to make progress on important topics. I also appreciate that you stated I was allowed to cry in your office once a year... and then that you stopped counting. Thank you to Pierre for engaging with detail. Ezra – you have given me new insight into how to think about my core questions. And Fiona you make it feel like home here. Thank you also to the broader community at MIT and beyond who made the experience more interesting, the struggles more bearable, and the small wins bigger.

Special thanks go to all those scientists around the world, and particularly those in Africa, who opened their labs to me and provided insight into their daily lives, without which none of this research would have taken place. This dissertation also benefited from generous funding from MIT's Legatum Center, the MIT-Africa Program, and a MIT Sloan PhD fellowship.

I couldn't have written this thesis without the support of my family and friends. Thank you to my siblings, Zoe, Jeremy, and the family we chose: Matt, Tessa and Anne, for helping me to see perspective, maintain a sense of humor and always paying for dinner. My in-laws, Tim and Kim, thank you for giving me a home here in the US. Tucker, you have been my rock. Without your commitment to me, and your unwavering belief that I could do it – this thesis probably wouldn't exist. Sorry it took so long!

Above all, thank you to my mother, and my late father. You have given me everything. You instilled in me a sense that I should leave the world in a better place than I came into it, and this thesis is my first attempt to fulfill this life-long mission.

For the African scientists who lost their lives during the 2014 ebola outbreak.

Contents

Chapter 1

1 Introduction

1.1 Background

This dissertation explores the role of international ties in the development of scientific capacity in low-income countries.

While it is well established that science and innovation are central drivers of economic growth (Schumpeter 1942; Solow 1957; Abramovitz 1986; Romer 1990; Jones 1995), the production of new knowledge is dominated by a handful of countries. Scientific productivity around the world is highly skewed, with over 60% of global scientific publications emanating from countries that are part of the Organizational for Economic Co-operation and Development (OECD). In contrast, some countries and regions of the world are lagging far behind in terms of scientific output. Despite being home to over one billion people, sub-Saharan Africa (excluding South Africa) accounts for less than 0.5% of global scientific publications, and not one person born or living in these countries has ever been awarded a Nobel prize in science. Per capita, OECD countries produce over 45 times the number of publications of African countries (Figure 1), and this skew is greater than that of other economic activities (Figure 2). That said, the difference in the relative scientific productivity between African and OECD countries has declined over the last 20 years (Figure 2).

One popular approach to thinking about the determinants of scientific and innovative activity in countries or regions of the world, is to examine country level institutions. The national systems of innovation (and related) approaches (Freeman 1987; Lundvall 1992; Nelson 1993; Furman and Hayes 2004; Hu and Mathews 2005) describes variation in innovative activity as a function of domestic institutions and the linkages between these institutions in a country or region. While this approach has made significant progress in describing a relationship between innovation and location, my central argument is that science is a global effort, and so an examination of domestic institutions could be insufficient. I consider science as a global system, with for the most part pooled inputs, and a shared audience around the world. This is particularly true for African country scientists – who have little to no domestic funding or outlets for science, and so rely on the rest of the world. One particular pooled input worth mentioning, which is shared even if countries do have adequate funding and outlets, is knowledge. If we think of scientists as standing on the shoulders of other scientists, global knowledge production is a crucial input into subsequent knowledge generation. However, knowledge doesn't travel very well over geographic distance and national borders, so for scientists and innovators located far from the predominant sources of production, it can be a challenge to access this input into the ideas production function.

Compounding the difficulties associated with geographic distance from dominant producers and users of knowledge, I argue that there is a global status ordering of scientists. Status can drive access to shared knowledge, resources and recognition (Merton 1968; Cole and Cole 1968; Zuckerman 1970, 1988; Allison et al 1982; Podolny 1993), therefore for those outside of exclusive circles, it can be a challenge to access crucial inputs to do frontier science. So, if location determines relative position in the global status ordering, and this drives access to knowledge, resources and recognition, this could account in part for the observed skew in scientific production around the world. With African scientists by many measures less elite than their global peers, this could present additional difficulties in improving their scientific performance.

One way to overcome the disadvantages of stratification that social or physical outsiders in a system may experience, is to develop relationships with insiders who can in theory share their advantage, either by sharing knowledge or resources, or transferring status (Blau 1964; Granovetter 1973; Goode 1978; Marsden 1983; Latour 1987; Burt 2010). In part due to declining communication costs, African scientists are increasingly developing relationships with scientists from high income countries, or insiders in global science. In fact, rates of international collaboration amongst African scientists are the highest in the world, with over 80% of African publications coauthored by extra-regional collaborators.

In this thesis I ask several inter-related questions surrounding the role of international relationships in the development of scientific capacity in sub-Saharan Africa. Each chapter helps to uncover a different feature of the way in which, and the scope by which, international ties impact African science, and ultimately facilitate technological catch-up and economic growth. I use empirical methods and novel data to ascertain for the most part a causal association between relationships and performance of African scientists, and combine this with qualitative data and a deep understanding of the context that I gathered over a total of nine months in the field.

1.2 Overview of Chapters

In the first chapter of the thesis I explore whether ties with international scientists impact the productivity of African scientists. Relationships with international scientists – who are more prominent by many measures – can be a central source of productivity and influence. Observational evidence from the scientific setting and beyond finds that ties with prominent affiliates have a positive impact on the performance of lower status actors. In practice though, relationships with those more elite are limited to high achieving or high potential individuals, making their causal value very hard to measure. Moreover, observational evidence that relies on samples of actors that have already formed and maintained such relationships cannot issue predictions about what the impact of such ties would be on those who don't yet have them. I address this measurement problem by examining the impact of an unexpected opportunity to build relationships with more prominent affiliates. The 2014 West African ebola epidemic afforded scientists working in endemic countries an unexpected opportunity to build relationships with more prominent affiliates from around the globe. I estimate the impact of the ebola epidemic on publication rates and international collaborations of 52 endemic country scientists by comparing the change in their outcomes before and after the epidemic with that of a matched sample of 250 similar scientists from non-endemic countries. I find evidence of a persistent post-epidemic boost in publication rates and international collaborations for endemic country scientists. However, these results are only found for those endemic country scientists who were already well connected with international scientists – with social capital – and working in disease areas similar to ebola – with intellectual capital – before the epidemic. This evidence highlights the importance of opportunities to build relationships with more prominent affiliates, but at the same time raises concerns over the potential implications of networks on inequality within groups outside the exclusive elite.

Driven by the observation that ties can create inequality amongst African scientists, in the second chapter of the thesis I assess whether one scientist's international ties can have a spillover impact on the performance of other African scientists. Prior research focusing on the ability of 'brokers' – actors that span networks – to share their networks and knowledge renders mixed findings. While in theory a broker could share networks and knowledge, in practice there are incentives not to share or borrow, and for challenges to arise. However, there are conditions under which we would expect such sharing to exist, namely: (a) when knowledge is codified and freely available; (b) when brokers have incentives to share; and (c) for 'outsiders' in a network who otherwise cannot access central resources and knowledge. One setting that meets these three conditions is returnee scientists moving back to African institutions following training in the United States forming what I call a 'core/periphery bridge'. I assess whether returning scientists who have developed international ties can share their knowledge and connections with non-migrants in the institutions they return to. Specifically, I study the effect of the return home of 112 HIV researchers trained in top universities in the United States under the National Institute of Health Fogarty AIDS International Training and Research Program between 1988-2014. I construct a panel dataset of 1,657 non-migrant African scientists who are affected by these return events in that they are working in related fields in the institution to which the American-trained scientist returns. I compare changes in publication outcomes of scientists working in institutions receiving an American-trained return migrant before and after the return event with those of observably similar scientists in African institutions not receiving a return migrant. The results reveal increases in the rate at which non-migrant scientists collaborate with scientists from the American training institution of the returning scientist following the return event. Non-migrants also increase citation rates to publications' of scientists based in the American training institution of the returning scientist. Furthermore, non-migrants experience a persistent increase in publication output following the arrival home of an American-trained scientist, particularly in HIV research. The effect is most pronounced for non-migrants who are not connected to OECD country scientists prior to the return event. The findings support the idea that a returning scientist forming a core/periphery bridge benefits periphery actors. In settings where 'outsiders' struggle to access knowledge and resources that are usually reserved for exclusive 'insiders', this kind of bridge in the network can help through providing legitimacy to the outsiders.

In order to ascertain whether these dynamics observed at the micro-level hold at a country level and contribute towards macro-level improvements in scientific capacity, the third chapter of the thesis assesses the association between international knowledge spillovers, cross country teams and African publication output. In this chapter I estimate the parameters of the ideas production function for African countries. I do so by considering international knowledge spillovers and cross-country teams as core determinants of technological catch-up, and estimate the elasticity of African publication output to foreign knowledge production. Using data of sub-Saharan African countries' scientific output between 1976 and 2016, I provide evidence for three main findings. First, the level of production of scientific output increases with the stock of ideas already discovered in a given country, as well as the level of human capital devoted to the scientific sector. Second, the level of production of scientific output is declining in the worldwide stock of ideas. That being said, the level of production of scientific output of African countries increases with the stock of ideas discovered in the ex-colonial power, as well as the levels of R&D funding of the ex-colonial power. This relationship is growing stronger over time, and is moderated by the size of the African country and their distance to the frontier. Third, the rate of collaboration between African and international scientists, particularly those from ex-colonial countries, is increasing over time. However, once this calendar trend is accounted for, international collaborations are more common for countries further behind the frontier. In an attempt to reconcile the findings, I find that the positive relationship between frontier country knowledge stocks and African publication output is moderated by the proportion of the African country scientific workforce that is involved in teamwork with the frontier country. I argue that international teamwork facilitates benefits from international knowledge spillovers and subsequent technological catch-up, particularly at earlier stages of development and for smaller countries. Overall, these findings are consistent with the concept that the rate of developing economy technological catch-up is associated with the production of knowledge in those developed countries with which they have relationships. Moreover, the findings suggest that knowledge, even that captured in scientific publications, is not easily accessible beyond these bilateral relationships, which has implications for programs and policies aiming to facilitate technological catch-up.

1.3 Conclusion and Future Directions

In conclusion, the ambition of this research is to better understand the drivers of innovation in emerging economies, and specifically how international ties can shape access to crucial inputs and thus facilitate technological catch-up. One important next step in this agenda is to understand how international ties shape the direction of innovation in low-income countries. While each of the essays point to a relationship between international ties and research direction, future research will explore how international relationships drive scientific specialization in low-income countries, and potential implications of these shifts in research direction. Another next step in this agenda is an exploration into the boundary conditions of the benefits of international ties, and the scope of activities that generate similar outcomes to those observed in the cases studied in this thesis.

In parallel to this line of research, two important lines of inquiry that I hope to pursue merit mentioning.

First, measuring emerging economy science and innovation is a challenge. Classic measurements of both innovation inputs and outputs designed for more advanced economies may be less applicable in this setting. Prior research on clinical trial infrastructure in Africa (Fry 2016) reveals that an index developed to predict clinical trial activity in more developed countries provides less explanatory power for African countries. Moreover, the use of patents and publication counts to measure innovation output in emerging economies presents an issue. Patenting is rare in this context, and coauthoring a publication with international collaborators can mean many things in terms of actual innovative capacity. Further study should design and test some more appropriate measurements that capture inputs as well as outputs, incorporating both brand new to the world innovation as well as a measure of innovations that have been adapted from elsewhere.

Second, it is crucial that we have a better understanding of the consequences of science and innovation activities for emerging economies. A core question of policy makers, business leaders and donors is what innovative activities to engage in (initially) in a given location. For emerging economies, these considerations must include competitive advantage, national priorities, existing strengths and what they can borrow from elsewhere. Future research should design and implement projects that can inform this question and subsequent decision making.

References

Abramovitz, M., 1986 *Catching Up, Forging Ahead and Falling Behind* Journal of Economic History 46: 385-406

Allison, P.D., Long, J.S, Krauze, T.K., 1982 *Cumulative Advantage and Inequality in Science* American Sociological Review 47 (5): 615-625

Blau, P.M., 1964 *Exchange and Power in Social Life* New York: Wiley

Burt, R.S., 2010 *Neighbor Networks: Competitive Advantage Local and Personal* Oxford University Press

Cole, J.R., Cole, S., 1973 *Social Stratification in Science* Chicago: University of Chicago Press

Freeman, C., 1987 *Technology Policy and Economic Performance: Lessons from Japan* Pinter, London, UK

Furman, J.L., Hayes, R., 2004 *Catching up or Standing Still? National innovative productivity among 'follower' countries, 1978-1999* Research Policy 33: 1329-1354

Goode, W.J., 1978 *The Celebration of Heroes: Prestige as a Social Control System* Berkeley, CA: University of California Press

Granovetter, M.S., 1973 *The Strength of Weak Ties* The American Journal of Sociology 78(6): 1360-1380

Hu, M-C, Mathews, J.A., 2005 *National Innovative Capacity in East Asia* Research Policy 34(9): 1322-1349

Jones, C., 1995 *R&D Based Models of Economic Growth* Journal of Political Economy 103: 739-784

Latour, B., 1987 *Science in Action* Cambridge MA: Harvard University Press

Lundvall, B.A., 1992 *National Systems of Innovation: Towards a Theory of Innovation and Interactive Learning* London, UK, Pinter

Marsden, P.V., 1983 *Restricted Access in Networks and Models of Power* American Journal of Sociology 88: 686-717

Nelson, R.R. (Ed.), 1993 *National Innovation Systems: A Comparative Analysis* Oxford University Press, New York, NY

Podolny, J.M., 1993 *A Status-Based Model of Market Competition* American journal of sociology 98(4): 829-872

Romer, P., 1990 *Endogenous Technological Change* Journal of Political Economy 98: S71-S102

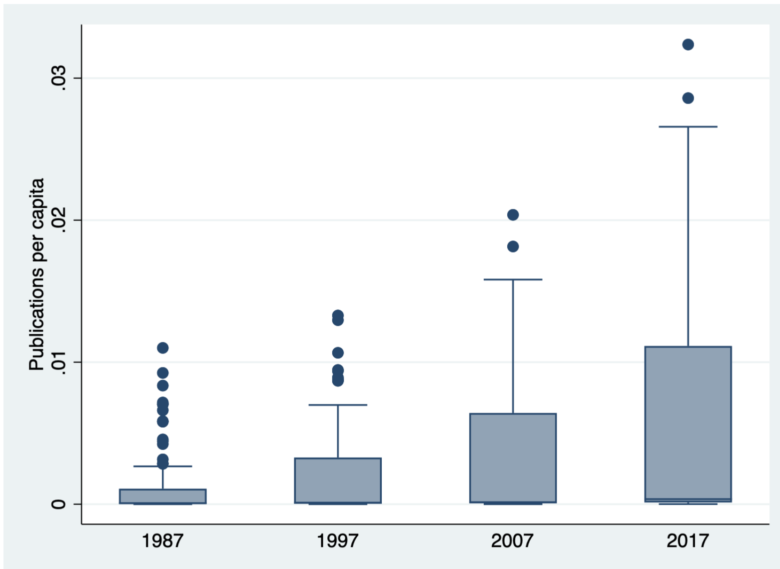
Schumpeter, J., 1942 *Capitalism, Socialism and Democracy* 2nd ed.George Allen & Unwin, Ltd., London

Solow, R.M., 1957 *Technical Change and the Aggregate Production Function* The Review of Economics and Statistics: 312-320

Zuckerman, H., 1970 *Stratification in American Science* pp 235-57 in E.O. Laumann (ed.). Social Stratification. Indianapolis: Bobbs-Merrill

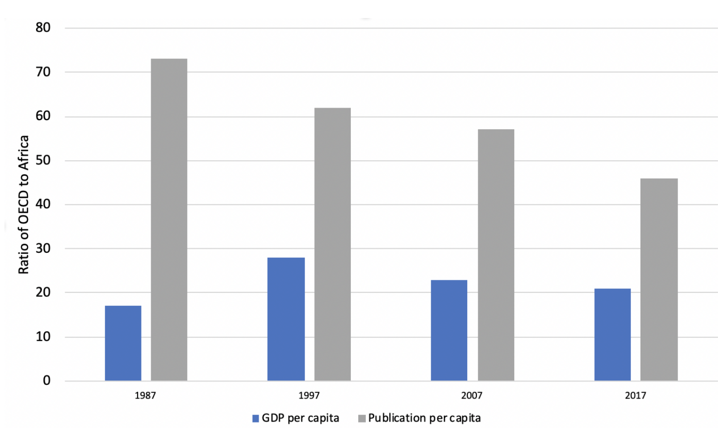
Zuckerman, H., 1988 *The Sociology of Science* In Handbook of Sociology, edited by N.J. Smelser. Newbury Park, CA: Sage

Figure 1: Skew of Publications Across OECD and African Publications



Note: Distribution of publications per capita across OECD and African countries across ten year periods.

Figure 2: Ratio of Publications and GDP Per Capita OECD:African Countries



Note: Ratio of publications per capita and GDP per capita of OECD to African countries across ten year periods.

Chapter 2

2 Viral Privilege: Evidence from the Ebola Epidemic

Abstract

Relationships with more prominent affiliates can be a central source of productivity and influence. In practice, relationships with those more elite are limited to high achieving or high potential individuals, making their value very hard to measure. I address this problem by examining the impact of an unexpected opportunity to build relationships with more prominent affiliates. The 2014 West African ebola epidemic afforded scientists working in endemic countries an unexpected opportunity to build relationships with more prominent affiliates from around the globe. I estimate the impact of the ebola epidemic on publication rates and international collaborations of endemic country scientists by comparing outcomes of endemic country scientists with outcomes of a matched sample of scientists from non-endemic countries before and after the outbreak. I find evidence of a persistent post-epidemic boost in publication rates and international collaborations for endemic country scientists. However, these results are only found for those endemic country scientists who were already well connected with international scientists and working in disease areas similar to ebola before the epidemic. This evidence highlights the importance of opportunities to build relationships with more prominent affiliates, but at the same time raises concerns over the potential implications of networks on inequality within groups outside the exclusive elite.

2.1 Introduction

'Never Let a Good Crisis Go To Waste'

Winston Churchill

Status orderings are a feature of economic and social life. The advantage of status, and in particular how status matters independently from quality in competitive outcomes, is of significant interest as it plays a role in generating and sustaining inequality. Where there is uncertainty about the underlying quality of a producer, market participants rely on status signals to make inferences about quality (Podolny 1993) or to co-ordinate actions (Correll et al 2017). Insofar as perceptions reinforce the initial signal, through the allocation of recognition or resources, status orderings self-perpetuate and the connection between actual quality and the status distribution will weaken.

One way that actual quality and status can decouple is that actors with higher status receive greater recognition, holding quality constant. Although challenging, scholars have made progress on finding causal evidence of this phenomenon in the real world (Simcoe and Waguespack 2011; Azoulay et al 2014; Kim and King 2014). A second way that the link between quality and status can loosen is through relationships. Relationships with more prominent affiliates that transfer status and serve as a channel for resources can have a real impact on performance. Past research has advanced our understanding of this feature of status hierarchies and documented a linkage between relationships with more prominent affiliates and performance (Long et al 1979; Stuart et al 1999; Burton et al 2002).

However, making progress on this line of research is difficult because of the link between relationships, quality, and performance in real-world settings. First, while researchers have documented that actors with a greater quantity and quality of relationships with higher-status actors have superior performance (Long et al 1979; Stuart et al 1999; Burton et al 2002), it is hard to disentangle whether that superior performance reflects differences in the likelihood of more able individuals to establish and maintain such relationships or whether such relationships directly contribute to performance itself. Second, there is good reason to doubt that just anyone can be placed in, and benefit from these kinds of relationships. Insofar as an actor's reputation is determined by the status of its affiliates, high status actors have a reason to avoid relationships with lower status actors due to the risk to their own reputation (Podolny 1993). It is possible that actors would need skills and know-how to be able to leverage a relationship with more prominent affiliates. Therefore, particularly in contexts where decisions are justified on meritocratic grounds, we might expect a limit to the benefit of such relationships.

To identify the impact of relationships with prominent affiliates on performance, it would be important to be able to observe an individual's relationships and performance, whereby relationship formation is independent of underlying quality. An ideal experiment would somehow allow some individuals (but not others) to establish relationships with more prominent affiliates and in so doing so the researcher could evaluate the performance of individuals both before and after the relationship is established, as compared to individuals not subject to such relationships.

The purpose of this paper is to implement this logic using an unanticipated ebola epidemic that affected some nations and not others, and provided scientists working in endemic countries an opportunity to build relationships with more prominent affiliates, whilst scientists in non-endemic but otherwise similar countries were left unaffected. The 2014 ebola epidemic in West Africa caused scientists from around the world to turn their research agendas to focus on ebola. With unique access to patient populations, local knowledge and a presence on the ground, endemic country scientists experienced an unprecedented opportunity to build relationships with more prominent scientists from around the globe. I compare changes in publication rates and international collaborations of endemic country scientists before and after the epidemic to those of observably similar scientists in non-endemic countries.

In a test of the proposition that an opportunity to build relationships with more prominent affiliates can improve performance, this paper finds that an opportunity can help, but that an average effect hides striking heterogeneity. I find that an individual's ability to leverage an opportunity to build relationships with more prominent affiliates is shaped by their intellectual and social capital. Specifically, I find that scientists in endemic countries with both prior tropical disease focus – intellectual capital – and international connections – social capital – experience large and persistent increases in publication rates and international collaborations following the ebola epidemic, whereas those without such intellectual or social capital experience negligible, or negative effects. Thus, the findings support the idea that relationships with more prominent affiliates can increase inequality amongst groups of less elite individuals. This additional layer of stratification amongst less elite groups has been overlooked in the literature and could go some way to explaining persistent stratification across a variety of social systems.

The rest of the paper proceeds as follows. Section 2 reviews the literature on status, relationships with prominent affiliates and performance. Section 3 presents the setting and describes the natural experiment exploited in the paper, the 2014 West African ebola epidemic. Section 4 describes the data, measures and statistical methods. Section 5 discusses the results. Section 6 concludes and outlines implications of the findings.

2.2 Theoretical Framework

Academic work building on the observation that status orderings are pervasive has explored the link between status and performance and its role in perpetuating inequality, arguing that (1) where there is uncertainty about the underlying quality of a producer, social positions influence beliefs about an actor, and (2) beliefs impact outcomes. An association between belief and outcomes can cause status orderings to self-perpetuate through the allocation of resources and attention, resulting in cumulative advantage (or disadvantage). Merton coined the phrase the 'rich get richer and the poor get poorer', sometimes referred to as the Matthew effect, to describe the phenomenon.[1]

The idea that social cues weaken the link between actual quality and status and reinforce unequal advantage has motivated two distinct streams of literature that aim to understand the mechanisms driving this uncoupling. In one type of status advantage a shift in status increases the attention and resources given to an actor, holding quality

[1] The term 'the Matthew effect' was coined by Robert Merton in 1968, who credited Harriet Zuckerman as a co-author of the concept, to describe inequality in the way scientists are recognized for their work. It is now applied to cumulative advantage of economic capital more generally. It takes its name from the parable of the talents in the Gospel of Matthew.

constant. Quasi-experimental studies find that status matters in the allocation of attention and recognition, and that this is mostly true for actors or products with which there is greater ex-ante uncertainty (Simcoe and Waguespack 2011; Azoulay et al 2014; Kim and King 2014). Moreover, status influences the allocation of resources (Sorenson and Waguespack 2006), which can reinforce the distribution through actual improvements in outcomes.

Another type of advantage associated with status is not a focal actor's own status, but rather the possibility that relationships with more prominent affiliates can influence outcomes. There are two possible mechanisms by which relationships with those more prominent can improve outcomes. First, prominent affiliates can share their knowledge and resources. It is well established that knowledge and resources flow through relationships (Granovetter 1973; Marsden 1983). Thus under the logic that prominent actors have superior access to knowledge and resources as a consequence of their position in the status ordering, this can be transferred to their less prominent affiliates. Second, prominent affiliates can act as 'sponsors' to those less elite (Blau 1964; Merton 1973; Goode 1978; Latour 1987; Podolny 1993; Burt 2010). A number of scholars have argued that perceptions of an actor's quality can be shaped by their affiliates, particularly when there is uncertainty surrounding the true quality – which is often the case for new or less distinguished actors. This implicit transfer of status can serve to signal the less elite's quality (Spence 1974), and subsequently influence their access to connections, knowledge, resources and attention. Regardless of the pathway, one would expect that actors with prominent affiliates have superior access to connections, knowledge, resources and attention that may lead to subsequent actual or perceived improvements in performance.

To build on these ideas, researchers have sought to document the link between relationships with prominent affiliates and performance. In a study on job placements of graduate students, Long et al (1979) find that the prestige of a doctoral department and mentor is correlated with the success of the placement of the graduate student in their first job. Beyond the scientific setting, Stuart et al (1999) measure relative outcomes for entrepreneurial firms that are affiliated with prominent partners, and find that entrepreneurial firms with more prominent associates go to initial public offering (IPO) faster than comparable firms without such prominent associates. Relatedly, Burton et al

(2002) measure outcomes for new ventures with more prominent prior employers, finding that new firms coming out of more prominent firms are more likely to pursue innovative strategies and to attract external financing.

Although this literature provides many insights into the link between relationships and performance, some issues remain unresolved. First, Stuart and Sorenson (2007) highlight the difficulty in attributing an individual's outcomes to their relationships. Second, the conditions under which a relationship with more prominent actors is most beneficial has not yet been explored.

Challenges in Measuring the Value of Relationships

The establishment of a causal link between relationships and performance is extremely difficult (Manski 1993; Mouw 2006). Researchers using the standard approach of measuring a link between relationships and performance of actors in observational data face three major challenges. First is the problem of unobserved heterogeneity. Even if the researcher carefully controls for individual level attributes, unobservable features of an individual could drive both the ability of that individual to form and maintain relationships with more prominent affiliates, as well as performance. If this takes place, researchers could conflate the impact of relationships on performance with underlying features of the individual. Second is the problem of reverse causality; superior performance could lead to relationship formation instead of the other way round. In this instance, researchers could be over-estimating the value of relationships. Third, a common problem in studies of this kind is selection on the dependent variable. An examination of outcomes of those who already have relationships makes it extremely difficult to both understand what happens to those who have relationships that do not survive to observation, and to define an accurate control group who are comparable on every dimension aside from having the relationships.

One approach to overcome these challenges is to manipulate relationships with more prominent affiliates, holding all else constant. In this study, I focus on evaluating the impact of an opportunity that is randomly presented to some individuals to build relationships with more prominent affiliates. Thus I am able to measure the causal impact of relationships with more prominent affiliates, allowing for attrition of the relationship, as well as allowing for the possibility that it is not activated into an observable relationship. This allows for a realistic estimate of the role of relationships with more prominent affiliates, and a test of the following hypothesis:

Hypothesis 1 (H1) *An opportunity to build relationships with more prominent affiliates improves performance.*

Limits to Opportunities to Build Relationships

There is an argument to be made that the effects of an opportunity to build relationships with those more prominent may differ across subsets of the population. This may not have been picked up in studies using observational evidence if opportunities to build relationships were unequally distributed in the first place across groups of actors, or if inclusion in study samples was dependent on having an observable relationship.

Relationships are costly to form and maintain (Burt 1995; Jackson et al 2008; Rivera et al 2010), and by the converse of the logic outlined above, high status actors have an incentive to avoid relationships with less elite actors as it could threaten their own status (Podolny 1993). It is plausible that even once given an opportunity to build a relationship, there are limits to the extent to which individuals are able to take advantage of it. Recent studies that estimate the causal impact of peers on student and entrepreneurial outcomes report variation in the effectiveness of randomly assigned ties (Carrell et al 2013; Koning 2016; Hasan and Koning 2019). The authors of these studies attribute any failure to benefit from randomly assigned peers to endogenous patterns of social interactions following randomization. In other words, even once provided with a relationship, there is variation in the extent to which individuals activate and leverage the relationship.

To investigate the heterogeneous effect of an opportunity to build relationships with more prominent affiliates, I focus on two attributes of an individual that could moderate the impact of an opportunity: intellectual and social capital.

Intellectual Capital. Relationships provide a way to access complementary knowledge and skills (Jones 2009). Scientists self-report that they form collaborations based on shared interests and complementary skills (Hara et al 2003), and inter-firm alliances are observed more often when information on the potential partner's capabilities and resources is available (Gulati 1999), or when the stock of knowledge of the potential partner is greater (Ahuja 2000). With the local nature of many forms of knowledge (Nelson and Winter 1982; Levinthal 1997), the potential partner's stock of knowledge relative to the problem the partnership is trying to solve is likely to affect the value they bring to the partnership. Thus, the likelihood that a relationship actually forms and endures, and the subsequent benefit that an opportunity to build relationships has on performance should be greater for those with more relevant intellectual capital.

Social Capital. The overlap in social networks of potential partners can reduce search and co-ordination costs of a new relationship. Potential partners within the same social network are likely to have more access to information on each other, and 'embedded' relationships within the same network are more likely to be reliable and benefits from shared norms (Gulati 1995; Ahuja 2000; see Stuart and Sorenson 2007 for a review of embedded exchange). In particular, new relationships within the same network are more likely to be perceived to be reliable as cohesive networks increase the likelihood of sanctions against individuals violating norms (Coleman 1988), and enable communication of reputation effects (Reagans and McEvily 2003). In settings with high uncertainty, such as the scientific setting, referrals and trust that come with embedded relationships are likely to be of particular importance in the formation and maintenance of relationships. That being said, the greater the overlap in social network – or the more relevant the social capital is – between a focal actor and more prominent affiliates, the more likely it is that the relationship forms and endures, and that any performance benefits arise from an opportunity to build relationships.

In arguing that relevant intellectual and social capital of an individual limit the benefit from an opportunity to build relationships with more prominent affiliates, I expect the following relationship to hold:

Hypothesis 2a (H2a) *Relevant intellectual and social capital moderates the positive impact of an opportunity to build relationships with more prominent affiliates.*

Given that H2a suggests limits to the impact of an opportunity to build relationships with more prominent affiliates, to the extent that those better positioned to leverage the opportunity are also those who are already better performing, inequality amongst groups of less elite is increasing in the presence of such an opportunity. One setting in which this idea has been explored is that of the impact of globalization on low-income countries. Recent models support macro-level evidence of the impact of globalization on inequality within low-income countries (Kremer and Maskin 2006; Maskin 2015). Kremer and Maskin (2006) propose a skills matching model to explain this phenomenon. Specifically, the model proposes that high-skilled workers in a low-income country are able to benefit from globalization as they collaborate with high-income countries, but those individuals who are low-skilled are not able to participate and thus are excluded from the benefits of globalization. This is summarized in the following quotation from Kremer and Maskin (2006):

> *'The key insight is that the globalization of the production process may benefit only those in the developing country with a skill level sufficiently close to that of their rich country collaborators, thus marginalizing low-skill workers in the developing country.'*

The prediction of this skill matching model is that globalization – or an opportunity to build relationships – increases inequality in low-income countries, and that those low-skilled prior to the opportunity are left behind. In a continuation of the argument above, I suggest that in addition to the level of skill, or intellectual capital, prior social capital determines who benefits from an opportunity to build relationships. Thus in a group of less elite individuals to the extent that individuals with relevant intellectual and social capital are also higher performing ex-ante, an opportunity to build relationships with more prominent affiliates will increase levels of within group inequality. This leads me to my final hypothesis:

Hypothesis 2b (H2b) *An opportunity to build relationships with more prominent affiliates increases inequality amongst less elite groups.*

The remainder of this paper tests these propositions through examining the impact of an unexpected opportunity to build relationships with more prominent affiliates. The next section provides details of the empirical setting and approach.

2.3 Setting and Empirical Approach

This study exploits a unique natural experiment that provided a group of scientists the opportunity to build relationships with more prominent affiliates. This research design differs from extant research in two main ways. First, the natural experiment employed provides a plausibly random allocation of an opportunity to some scientists to build relationships with more prominent affiliates. Second, I create a sample of matched control scientists who did not receive the opportunity to build relationships with more prominent affiliates, but in every other way are observably similar prior to the event. This results in a sample consisting of scientists who are subject to the opportunity and those who are not with which to assess whether the opportunity to build relationships with more prominent affiliates affects performance. I analyze the change in performance of an individual after the opportunity, compared to that of a control scientist, in a difference-in-differences framework to avoid bias due to any constant, unobserved differences between scientists. The remainder of this section describes the setting and the natural experiment studied, incorporating a discussion of the broad empirical approach. The following section describes details of the data and statistical methods used.

West and Central African Scientists

The setting for the empirical work is academic scientists in West and Central Africa. The study's focus on this setting can be justified on substantive grounds. A long history of academic work has explored the sciences as a context for studying the effects of status (Merton 1968; Cole and Cole 1968; Zuckerman 1970, 1988; Allison et al 1982). While many of these studies have focused on status orderings of scientists within a country or region, in many respects science is a global community, with a global status ordering. Scientists around the world move freely between nations, compete for the same journal space and resources, and seek recognition and reward from the same gatekeepers. Within academic sciences, institutional affiliation is a primary determinant of status of an individual. With no university in West or Central Africa ever ranked in the top 800 research universities in the world,[2] as well as no Nobel prizes in science ever awarded to scientists from the region, by many measures scientists in West and Central Africa can be considered less elite than scientists from other regions of the world. The position of West and Central African scientists in a global status ordering is particularly visible when comparing to scientists based in OECD countries. OECD countries are home to the world's elite research institutions, and scientists located in these countries account for well over 60% of global publications and possess the majority of global resources for science.

As well as being arguably less distinguished than their global counterparts, scientists from the region rely heavily on international connections to access resources necessary for scientific production. High quality training for scientists in the region is scarce, labs are poorly equipped, and domestic funding for science is negligible. In a survey of around 500 West African scientists carried out by the author in 2017, it was discovered that just under 50% of respondents carried out their graduate studies in Organization for Economic Cooperation and Development (OECD) countries, and that the predominant funders of science in the region are American and European funders, including the Wellcome Trust, the US National Institutes of Health (NIH), the Bill and Melinda Gates Foundation, and the European and Developing Country Clinical Trials Partnership (EDCTP).

Given these contextual factors, any change in West or Central African scientists' opportunities to build relationships with scientists from elsewhere in the world, particularly from OECD countries, provides the ideal conditions under which we would expect the hypotheses outlined above to hold. The remainder of this paper focuses on one change in the opportunity to build relationships with more prominent affiliates: the 2014 West African ebola epidemic.

The 2014 West African Ebola Outbreak

Ebola virus disease is characterized by severe and mostly fatal outcomes. The disease is spread through direct contact with infected people or animals and tends to affect populations in sudden increases in incidences over a short space of time and space in the form

[2] https://www.usnews.com/education/best-global-universities/articles/methodology accessed on 3.5.20

of an outbreak. With no approved vaccination or treatment for the disease, outbreaks are catastrophic. Following its discovery in 1976 in the country known then as Zaire, there have been around twenty outbreaks throughout Africa. In March 2014 the World Health Organization (WHO) reported the first cases of an ebola outbreak in Guinea, West Africa and by August 2014 the WHO had declared ebola a Public Health Emergency of International Concern. Two years later, when the last case was confirmed in 2016, the virus had spread to ten other countries, but was concentrated in three countries at the epicenter: Guinea, Liberia and Sierra Leone. Over 30,000 cases and 11,000 deaths were attributed to the disease in these three countries (Figure 1). Experts associated the devastating impact to population movement across the porous borders between the countries. The epidemic had been the world's largest and deadliest in recorded history, with a WHO statement in 2014 recognizing that *'the ebola epidemic ravaging parts of West Africa is the most severe acute public health emergency seen in modern times.'*

The nature of the symptoms and spread of the disease captured the attention of the globe's media, policy makers and donors and the world watched as the epidemic in West Africa intensified. While public health interventions scaled, science was being touted as one of the potential solutions to the ghastly epidemic. With very little published research on the disease, a better understanding of the virus, its spread and mutations, and development and testing potential vaccinations and cures was one of the more hopeful avenues to contain the epidemic.

Scientists around the world turned their research agendas to focus on the disease (Mutters et al 2018) and global funders increased resources available. An estimated USD $435 million was spent in 2014 and 2015 on ebola alone (Fitchett et al 2016), the majority of which came from OECD countries, and particularly from the US government (Moran et al 2014). A number of foreign scientists took an interest in, started projects on, and visited, the endemic countries. The following quotation from a West African scientist illustrates the level of engagement of the international research community and the sudden attention given to endemic countries:

> *'During ebola a lot of scientists came in. It was quite a chaotic environment. Ebola was exciting for the research community. It is dangerous, little research*

had been done, there is no approved treatment. Ebola had all the right reasons to attract international researchers to Sierra Leone.'

Many of the foreign scientists researching the ebola outbreak worked together with local researchers and made use of pre-existing institutions in order to efficiently gain access to populations, gather samples,[3] obtain local knowledge and streamline their own research process. The local scientists who were well embedded in the hospitals and laboratories as well as the government (which can have complex research clearance procedures) were in high demand as collaborators to the foreign researchers. A small group of scientists in the endemic countries suddenly found themselves at the epicenter of the subject of some of the most topical research at the time. Two West African scientists confirmed their value as collaborators during the epidemic:

'You cant just go to a place to conduct research. They [the foreign researchers] were trying to attract locals therefore.'

'You want someone on the ground who can help you to achieve your aim and make things happen.'

This international interest in the outbreak and affected countries generated new relationships between endemic countries and global scientists and afforded scientists in endemic countries an opportunity to build relationships with more prominent affiliates from around the world.

Control Scientists

It might be reasonable to expect that a comparison of the outcomes of endemic country scientists before and after the epidemic, i.e. before and after they were provided with the opportunity to build relationships with more prominent affiliates, would give a causal estimate of the impact of the opportunity. However, there is the concern that career age

[3] These samples were frequently sent back to the researcher's home country for analysis.

trends as well as general improvements in regional capacity could conflate the role that relationships play with improvements that may have occurred absent the opportunity. To alleviate this concern, I use a control group that consists of carefully matched group of scientists from non-endemic, but otherwise similar, countries within West and Central Africa. The inclusion of these control scientists in the empirical framework allows to account for underlying trends in career age, field and general regional changes. The next section describes how the treated and control scientists are selected and the statistical framework by which the impact of the epidemic is estimated.

2.4 Data and Statistical Estimation

Data

In order to measure the impact of the ebola epidemic on endemic country scientists I generate a sample of endemic and comparable non-endemic country scientists who are actively publishing at the time of the epidemic. Each scientist in the sample is linked to their full publication history and their collaboration patterns and publication rates traced year to year.

The challenges in generating a sample of scientists in a particular location and linking this with their full publication record are considerable. First, generating a scientist level publication record is complicated by the fact that scientists may have common names (for example, Smith J). Therefore it can be difficult to determine which Smith J published which paper, or a single scientist may go by more than one version of a name. Second, knowing where scientists are located in the absence of administrative or resume data is difficult. Fortunately, the first issue is resolved using the Elsevier Scopus publication database's author identifier, which is a unique identification number for each author contained in the database. The author identifier is developed using an algorithm that incorporates scientist name, coauthors and topic type and allows for scientists to change affiliations across publications. This identifier allows me to track publications for every researcher captured in the database. The second issue is resolved using author affiliation data in the Elsevier Scopus database. To ensure that a scientist's affiliation in a given publication represents their actual location and not a visiting appointment or remote affiliation I use a rule of thumb – if a scientist classifies her affiliation as being in a certain country in over 75% of her publications over a four year period, she is considered in this database as being based in that country in that time period.

I extract scientists based in West and Central African countries in 2010-2013 according to the procedure described above. I exclude scientists who have not published in biomedical or social sciences (assuming that engineers, for example, are less likely to be impacted by the epidemic). I also exclude scientists who are never first or last author in the four years prior to the outbreak (to exclude technicians) and those who stop publishing before 2013 (to exclude those who retired/deceased/moved before the outbreak). From this sample of 6,758 West and Central African scientists (Table 1), I identify the 61 'treated' scientists based in endemic countries: Sierra Leone, Liberia and Guinea.

To construct the control group I select a suitable sample from the entire set of scientists located in the other nineteen countries across West and Central Africa (Table 1). The control scientists are chosen using a coarsened exact matching procedure (Iacus et al 2011) so that their average career age, productivity, research area, rate of international collaborations and country level variables such as GDP per capita and number of scientists mirrors that of the treated scientists (see Appendix A for more details on the construction of the control group). At least one match (and up to 18 matches for each treated scientist) is found for 52 (85%) treated scientists, giving 250 control scientists based in non-endemic countries (Table 2).

Each treated and control scientist is linked to its full publication history and year to year activity traced. The final estimation sample includes observations for each treated and control scientist 4 years before and 6 years after the epidemic. The result is a balanced panel dataset with 3,020 scientist-year observations. In addition to publication data, I conducted 35 interviews with scientists from both West Africa and OECD institutions in July-August 2018, as well as site visits to both treated and control country sites. The interviews ranged from 1 to 2 hours, with site visits ranging from a half day to a week.

Measurement

I conduct two main analyses of scientist's performance. In the first analysis, the dependent variables are centered around measurements of publication rates, as is standard in studies on scientists' performance. In the second part of the analysis, the measures focus on international collaborations. It is possible that an opportunity to build relationships results in collaborative publications with more prominent, international scientists. As past studies have identified evidence of the connection between international collaborations and research impact (Van Raan 1998; Wagner and Leydersdorff 2005; Jonkers and Tjissen 2008), this is an important outcome to measure in its own right. A description of how the variables are generated is provided below.

Publication Rates Measures corresponding to the rate of publication include the number of publications in an observation year that a scientist is an author on, and an additional measure weighting each publication by its journal impact factor (JIF) – a measure of the frequency with which the average article in a journal has been cited in a particular year. These count measures have been used widely in previous studies of scientific productivity (Ding et al 2010; Azoulay et al 2010). Additional analysis using variations of measures, such as scientist role on projects, as well as topic choice are provided in Appendix C.

International Collaborations Measures corresponding to the rate of international collaborations of a focal scientist in an observation year are generated by extracting coauthor names and affiliations from the focal scientist's publication record. I focus on collaborations with OECD based scientists. With 63% of the global count of publications in 2013 containing authors affiliated with countries that are part of the OECD, the scientific ecosystems in OECD countries are regarded to be the central locations for the majority of scientific research. Collaborations with OECD based scientists are measured in two main ways: (1) the number of publications in an observation year with at least one OECD coauthor, and (2) the number of new OECD coauthors (i.e. OECD based scientists that the focal scientist had not previously coauthored with) in an observation year.

Descriptive Statistics

The descriptive statistics in Table 3 pertain to the set of 52 + 250 matched treated and control scientists. The covariates of interest are measured at baseline, just prior to the epidemic (end of 2013). A number of the covariates are balanced by construction, due to the coarsened exact matching procedure, for instance, the career age and collaboration patterns of scientists, but I also find balance of other key covariates that is not guaranteed by the matching procedure. Two features of the sample of scientists are worth pointing out. Around half of the sample scientists collaborate with OECD scientists in the year prior to the outbreak, although the median number of new OECD coauthors is zero. This suggests that although some scientists are connected with more prominent affiliates, the flow of new relationships is actually quite minimal at baseline. Interestingly, less than than 8% of endemic country scientists had experience in viral hemorrhagic disease research, which is a family of viral diseases including ebola. A broader categorization of research topic that incorporates ebola is neglected tropical diseases. Almost 30% of scientists published in neglected tropical disease areas in the year prior to the outbreak.

Statistical Estimation

In order to identify the effect of the ebola epidemic on endemic country scientists, I compare an endemic country scientist's outcomes after the epidemic relative to before, using a scientist fixed effect specification. The estimating equation (equation 1) relates endemic country scientist i's outcomes in year t to the epidemic.

$$\begin{aligned} E\left[y_{it}|X_{it}\right] = exp\Big[&\beta_0 + \beta_1 AFTER\_EPIDEMIC_t \times ENDEMIC\_COUNTRY_i \\ &+ f(AGE_{it}) + \delta_t + \gamma_i\Big] \end{aligned} \tag{1}$$

Where y is the outcome measure, AFTER_EPIDEMIC denotes an indicator variable that switches to one the year the ebola epidemic began (2014).[4] ENDEMIC_COUNTRY denotes an indicator for if the scientist is affiliated with an institution in an ebola endemic

[4]The post treatment variable considers the years 2014-2019 in the data as post epidemic. Generally

country. f(age) corresponds to a flexible function of the scientist's career age[5] as is standard to include in studies of scientist productivity (Levin and Stephan 1991) and γ_t stands for a full set of calendar year indicator variables to account for the fact that aggregate research activities may vary over time. δ_i correspond to scientist fixed effects, consistent with my approach to analyze changes in the scientist's output following the epidemic. Standard errors are clustered at the level of the individual scientist.[6]

The majority of the dependent variables of interest are skewed and non-negative (Figure 2 illustrates the distribution of publications prior to the epidemic (2010-2013 inclusive)). Due to the large number of zero's in the dataset, and following tradition in the study of scientific and technical change, I mostly present quasi-maximum likelihood (QML) estimates based on the fixed-effects Poisson model developed by Hausman et al (1984) (Appendix B provides estimates based on the fixed-effects Ordinary Least Squares models with (i) inverse hyperbolic sine transformation of the dependent variables and (ii) log transformation of the dependent variables).

2.5 Results

As a preliminary step I analyze whether a scientist's location at the time of the ebola epidemic influences the rate of ebola publications coauthored with OECD based scientists. Figure 3 illustrates average publication rates in ebola related research with OECD collaborators for endemic and non-endemic country sample scientists between 2010 and 2019.

there is a lag period of around a year between when scientific work is carried out and when the publication appears in a journal. However, interviews uncovered that during the ebola outbreak journals fast-tracked the relevant publications and published them in real time. Some scientists even noted that they worked with the journals to provide updates as the research went on. For this reason, the year that the epidemic started, 2014, is considered the first year when an effect could be seen.

[5]Data on actual career age is not available and so I deduce career age as the years passed since the first observable publication of a focal scientist as found in the Elsevier Scopus publication database.

[6]A threat to the empirical design exists if scientist's standard errors are correlated within a country due to the existence of country-level time-varying unobservables (Donald and Lang 2007). However, with just 10 countries, clustering the standard errors at the country level would not meet the asymptotic convergence assumption. Moreover, clustering at this aggregate level can lead to incorrectly estimated standard errors that are upwardly biased (Abadie et al 2017). Thus, given the fact that the sampling strategy does not exclude observations from the population that could be similar, I cluster the standard errors at the level of the individual.

Endemic country scientists experience a sharp increase in these publications following the start of the epidemic in 2014, while the rate of these publications for non-endemic country scientists remains relatively flat. This evidence aligns with the qualitative evidence on the increase in visibility and attention from the international scientific community, particularly OECD based scientists, that endemic country scientists experienced during the outbreak, and the rest of this section explores the impact of that opportunity to build relationships with more prominent affiliates on performance outcomes.

The Effect of an Opportunity to Build Relationships with More Prominent Affiliates (H1)

Figure 4 reports the trend in publication output and collaboration rates for endemic country scientists, and matched control scientists from non-endemic countries. The figure provides descriptive support for hypothesis 1, which states that an opportunity to build relationships with more prominent affiliates improves performance, measured as publication rates and international collaborations. Following the epidemic began in 2014, the publication and collaboration outcomes of endemic country scientists are consistently higher than that of non-endemic country scientists. Interestingly, comparing panels (a) and (b) suggests that this performance improvement was mostly in the quality of publications, evidenced through the relative increase in journal impact factor weighted publications, as compared to simple counts of publications. However, this raw comparison does not include any adjustment for career age or calendar year, and could be driven by differences in the baseline of scientists. I control for these differences using the scientist fixed effect treatment model described above.

Table 4 presents the core results estimating the specification presented in equation 1. Consistent with the raw data, I find support for H1, that an opportunity to build relationships impacts publication rates (columns 1 and 2) and international collaborations (columns 3a-b and 4), as indicated by the estimates for AFTER_EPIDEMIC × ENDEMIC_COUNTRY being positive and mostly significant across outcomes measured. I find a sizable 22% increase in the annual number of publications of an endemic country scientist following the epidemic (column 1), as compared to a non-endemic country scientist. Although the coefficient is not significant owing to the large amount of variation across scientist years, this is a relatively large increase in the number of publications. Given the baseline publication rate prior to the epidemic of around 0.73 publications per scientist per year, this would equate to around one additional publication per scientist in the six year period following the epidemic. To verify that this increase isn't driven by an increase in publications in low quality journals, column 2 measures the change in publications weighted by their journal impact factor. The even greater statistically significant 68% increase for the treated group as compared to the control group suggests that scientists mostly increase the quality of their publications following the epidemic. The results show that endemic country scientists publish with OECD based coauthors at a rate relatively greater than control scientists following the epidemic (column 3a). A rate that isn't matched by an increase in non-OECD coauthored publications (column 3b). Moreover, I find that following the epidemic, endemic country scientists form more than 120% more new OECD coauthors a year as compared to non-endemic country scientists (column 4). This amounts to around 2.5 additional new collaborations per year, which in the six year period following the epidemic gives over 15 additional collaborations formed. Appendix C provides additional results measuring changes in endemic country scientist's project role and research topic as a result of the epidemic.

The Moderating Effect of Intellectual and Social Capital (H2a)

The results up until this point show an average effect of the opportunity to build relationships with more prominent affiliates. The subsequent analysis examines whether the magnitude of the impact of the epidemic correlates with pre-epidemic attributes of the scientists.

Table 5 evaluates the moderating effect of intellectual capital on an opportunity to build relationships with more prominent affiliates. As a proxy for a scientist's relevant intellectual capital prior to the epidemic I construct a dummy variable that takes a value of 1 if the scientist has any publication in neglected tropical disease related research in the four years prior to the epidemic. Ebola is categorized as a neglected tropical disease, and given that very few researchers around the globe had experience in ebola research prior to the 2014 outbreak, many scientists turned their agendas from other neglected tropical diseases (such as malaria) to ebola. Thus in this definition of intellectual capital I assume that the knowledge required to research neglected tropical diseases is more transferrable to ebola research than the knowledge required to research, say, diabetes or cancer. The median likelihood of a sample scientist publishing in neglected tropical diseases prior to the epidemic is 0, and so this gives roughly similar sizes of sub-groups.

I split the sample of treated and control scientists into those who have a publication record in neglected tropical diseases prior to the epidemic, and the rest of the sample, and run the same specification on the two samples separately for selected outcomes. The results in the table illustrate that those endemic country scientists with a publication record in tropical diseases benefit the most from the epidemic. Moreover, the results suggest that intellectual capital predominantly influences quality (JIF-weighted publications) of publications, as well as the quantity of new OECD collaborators. Not only do those with relevant intellectual capital experience a greater positive impact from the epidemic, but those without such experience in neglected tropical disease related research – in columns 2, 4, and 6 – experience a decline in their performance following the epidemic, as compared to comparable non-endemic country scientists. During the epidemic, scientists who weren't well positioned to participate in the epidemic science with foreign scientists reported that they were compelled to stop their research projects due to safety concerns. Thus I interpret this as a lack of a positive impact of the opportunity, and the negative coefficients are a feature of the specific event studied and not necessarily generalizable beyond this setting.

Table 6 provides evidence of the moderating effect of social capital on an opportunity to build relationships with more prominent affiliates. As a proxy for a scientist's relevant social capital, I generate a dummy variable indicating whether the scientist has all of their publications coauthored with OECD based scientists in the four years prior to the epidemic. In doing so I assume that embeddedness with any OECD based scientists implies a sharing of norms and communication patterns with other OECD based scientists, an assumption that was verified during interviews. I choose this definition that requires a scientist to have published all of their publications with OECD based scientists in order to distinguish between those who are truly embedded, and those who have participated in research projects but may not be as familiar with norms and routines associated with the

OECD network of scientists. Almost all scientists have coauthored with OECD scientists at some point and the median proportion of OECD coauthored publications across scientists in the four years prior to the epidemic is 1. However, field work highlighted variation in the approach of West African scientists to collaborative field work. Some scientists consider themselves as working in the same lab (albeit remotely) as OECD based scientists, sharing equipment and funding with frequent visits between labs. These scientists tend to publish all of their publications together, while other scientists have more ad-hoc collaborative relationships with OECD based scientists. It is the former category of scientists – which encompasses 63% of the study sample – who have embedded relationships with OECD based scientists that I consider having relevant social capital during the epidemic.

Again, I split the sample into those who have significant experience publishing with OECD coauthors, and those who don't, prior to the epidemic. Table 6 illustrates that those endemic country scientists with a significant OECD collaboration record experience a greater positive effect of the epidemic than those without, although the benefit is much smaller than that of the intellectual capital (and not statistically significant in the difference between those with social capital and without). However, qualitatively there is a difference. The results suggest that embeddedness in the OECD scientific network predominantly influences the number of new OECD collaborators formed during the epidemic. While this could be due to a variety of factors, one possibility is that OECD connections can provide a referral, or endorsement, for new relationships to form. One scientist based in the United States describes how her coauthored publications with her West African collaborator prior to the epidemic played a role in the connections her African collaborator made during the epidemic:

> *'I got a lot of calls from people at medical schools in the US who wanted to be involved because they thought it [ebola research] was cool... I think I got more calls than [West African collaborator]. People saw that I had the American looking name so (thought) I must be in charge.'*

Not only is there suggestive evidence that foreign scientists search for African collaborators through leveraging the existing network, but West African researchers also used their pre-existing network to filter through requests for partnerships during the outbreak. Several West African scientists reported that they received many unsolicited requests for collaborations from foreign scientists during the outbreak. The scientists replied to requests from scientists who were known by their prior collaborators. For example, one endemic country scientist spoke about how s/he turned down a request from an 'out-of-network' foreign scientist *'one contacted me from [foreign country], he was evidently a rogue scientist who wanted ebola samples.'* Interestingly, the results in table 6 suggest that social capital alone has a limited impact on both the quantity and quality of output.

Finally, I explore whether intellectual and social capital are complementary forms of capital, or substitutable. Is it enough to just have one or the other, or do scientists need both? To understand the dynamics between different forms of capital I split the sample into four mutually exclusive groups: (i) scientists with neglected disease experience **and** a significant OECD coauthor record prior to the epidemic; (ii) scientists with a record in neglected tropical disease research **but without** always having coauthored with OECD based scientists: (iii) scientists who have coauthored all of their publications with OECD based scientists, but with **no** record in neglected tropical disease research: and (iv) scientists with **neither** neglected disease experience **nor** having coauthored all of their publications with OECD based scientists. I run the same specification on each sub-sample in Table 7 for selected outcomes. I find that scientists with both experience in neglected tropical disease research and OECD collaborations are able to benefit the most from the opportunity to build relationships (columns 1 and 4). This suggests that intellectual and social capital are complements to one another, consistent with the idea that it getting ahead is 'what you know and who you know'. Scientists with a publication record in neglected tropical diseases, but without significant experience publishing with OECD based collaborators also benefit from the epidemic, albeit to a much lesser extent than those with both types of capital. The outcome that suffers the most from not having relevant social capital is the number of new OECD coauthors. This is not a surprise if we expect that new coauthors are mostly formed through a referral system. On the other hand, scientists without a publication record in neglected tropical diseases, regardless of whether they have experience publishing with OECD scientists don't benefit at all from the epidemic (columns 3, 4 and 7, 8). This last result suggests that the minimum and sufficient criteria to benefit from an opportunity to build relationships with more prominent affiliates is relevant intellectual capital, while relevant social capital is a less binding constraint. However, having both intellectual and social capital is optimal.

To summarize, the impact of an opportunity to build relationships with more prominent affiliates on performance, as measured by publication rates and international collaborations, is positive and significant. The impact is greater for individuals with more relevant intellectual and social capital. While intellectual capital appears to be more important than social capital in driving improvements in both quantity and quality of output, the best position to be in in order to benefit from an opportunity to build relationships is to possess both forms of capital. This implies that programs and policies supporting democratization through network based interventions may not have the same impact for less well positioned groups individuals, and that additional support should be provided to balance the disadvantages of stratification.

The Effect of an Opportunity of Within Group Inequality (H2b)

Tables 5, 6 and 7 all illustrate increasing inequality within endemic countries following the epidemic. Not only do the tables provide evidence that those with prior tropical disease research experience and significant experience publishing with OECD based scientists experience the greatest positive impact of the epidemic, they also show that those endemic country scientists without such experience are either not at all, or negatively affected by the epidemic. Given the relative advantages that already exist for West and Central African scientists who are connected with OECD based scientists described in section 3, this additional ability to leverage the opportunity further exacerbates inequality amongst endemic country scientists. Whether this inequality smooths out or intensifies over a longer time period than the study period is an important question for future research.

Mechanisms

The mechanism(s) driving the observed performance effect is explored in interviews and site visits. In section 2 I propose two possible channels that could drive any observed performance benefits arising from an opportunity to build relationships with more prominent affiliates. First, more prominent affiliates can share their knowledge and resources. Second, more prominent actors can transfer their status to their affiliates. This sponsorship can signal quality under conditions of uncertainty, which can subsequently improve access to connections, knowledge, resources and attention. While these are very difficult to separate empirically (particularly as the latter results in the same measurable outcomes as the first), I find qualitative evidence supporting the existence of both channels during the ebola outbreak.

Numerous examples support the idea that endemic country scientists had new access to resources during the outbreak, obtained through their relationships with the global scientific community. For example, the team at the Broad Institute in the United States and their West African collaborators negotiated the donation of three Illumina Miseq genomic sequencers (each valued at around USD $100,000) to West African laboratories during the outbreak. Two West African scientists confirmed that the opportunity provided them with equipment, skills and funding more generally, stating:

> *'We got software by projects to help with analysis, and data collection equipment... we were given a projector.'*

> *'There was all this funding around capacity building and health systems strengthening that people benefited from.'*

Endemic country scientists also had new access to knowledge through their global relationships. One endemic country scientist described the process of observing and participating in the research process with the international scientists that came into the country during the outbreak, and summarized the new skills and knowledge gained through this process by saying: *'During ebola we learnt how to do research.... it's not rocket science.'*

Although harder to observe, there is suggestive evidence that the status of the global scientists who engaged in the ebola epidemic science with West African scientists played a part in endemic country scientist's subsequent access to resources and attention. Endemic country scientists, some for the first time, applied for international research grants, and targeted publications in top international journals with their new collaborators that they generated during the epidemic. Consistent with evidence on the importance of reputation in accessing crucial resources in science (Simcoe and Waguespack 2011; Azoulay et al 2014), a West African scientist recounted the value of their new collaborator's reputations in accessing crucial resources:

> *'I feel that the connection with international partners is necessary. Because they have the reputation to get funding...'*

Longer Term Impacts

Figure 4 provides descriptive evidence that the boosts in publication rates and international collaborations experienced during the ebola epidemic for endemic country scientists are sustained through 2019. Outcomes for endemic country scientists are consistently higher than those of matched, non-endemic scientists. I explore the persistence of these effects further in Table 8. Using a reduced sample of endemic and non-endemic country sample scientists who publish with OECD coauthors in the two years following the start of the epidemic, I run cross sectional logit regressions to estimate average differences in outcomes of these scientists in the years 2016-2019 (after the epidemic ended) between endemic country and control country scientists. Using this later time period I'm able to discern whether the effects were limited to the ebola time period (when there was, understandably) significant focus on ebola-related publications, or whether there are longer lasting effects. I find limited evidence that endemic country scientists are significantly more likely to have a publication in general (column 1). However, examining the role of scientists on projects, as proxied through their position on the publication, tells a different story. Although endemic country scientists are no more likely to be the first author on publications (a proxy for the role on projects as predominant data collector) (column 2a), or middle author (often technician roles) (column 2b), they are significantly more likely to be the last author on publications following the epidemic (column 2c). The coefficients can be interpreted as log odds, implying that endemic country scientists are around 2.4 times more likely to be the last author on a publication in a given year between 2016-2019. This suggests longer term positive consequences of the epidemic on endemic country scientists in terms of intellectual contribution to products, and the ability to secure funding. Furthermore, endemic country scientists are more than twice as likely to have a 'hit' publication than non-endemic country scientists in a given year (Table 8 column 3), which I define as having a publication in a journal that is within the 95th percentile of the citation distribution amongst fields that the sample scientists are publishing in. A West African scientist confirmed these longer term impacts of the epidemic on his career:

> *'My ebola papers took me to bigger journals. I had not published in the Lancet before. It was ebola that took me to the Lancet... I am an editor for PlosOne now because of my publications (during ebola)'.*

These longer term results are striking, and imply that relationships with prominent affiliates can help scientists to move up the skill/hierarchy ladder within project teams.

Alternative Mechanisms and Robustness Checks

I conduct several alternative estimations to test the robustness of the results. First I consider an alternative explanation that one could be concerned about driving the results. Second I probe the robustness of the experimental approach.

We might be concerned that the endemic countries also increased their domestic investment in science around the time of the epidemic, and so the observed effect could be in part driven by this, biasing the result of the impact of the opportunity to build relationships with more prominent affiliates. To alleviate this concern I explore two sources of data on science and R&D funding in West and Central African countries. First, I examine the Policy Cures online database on neglected disease R&D funding. The database provides results from a comprehensive annual survey of global funders, including governments, on the levels of spending into neglected disease R&D. None of the endemic countries feature in the database between 2007 and 2017 and I interpret this as preliminary evidence that levels of domestic investment in R&D are not obviously increasing from these endemic countries as a result of the epidemic. Second, I examine the UNESCO Institute for Statistics database on gross domestic expenditure on R&D per year. Again – none of the endemic countries have any data between 2007 and 2018, and so levels again are not obviously increasing from these endemic countries. Numerous publicly available reports on the state of African science and field work observation confirms that the levels of R&D spending by African governments are extremely low, or negligible, and have not changed significantly in the last 10 years. I therefore find it difficult to believe that the observed effect is being driven by a change in domestic R&D spending in endemic countries.

I probe the robustness of the main results, that endemic country scientists experience increases in publication outcomes following the epidemic, in a variety of difference-in-differences regressions with the outcome of JIF weighted publications presented in Table 9. First, I estimate a placebo experiment using a placebo epidemic year for the full sample. Using just the pre-epidemic data I use an event date of 2010, and estimate the before after impact for endemic country scientists. This tests whether endemic country scientists outcomes were improving prior to the epidemic, which, if true, could bias my results upwards. Reassuringly, the effect of being in an endemic country in a placebo epidemic year is smaller than the benchmark and statistically insignificant (column 2). Next I examine the possibility that the result is being driven by a handful of more productive scientists. I estimate the specification without the inclusion of scientists who have more than 5 publications in the four years before the epidemic – which is the 90th percentile of all sample scientists (column 3). This additional specifications does not change the findings of the main models.

2.6 Discussion

Benefits arising from relationships with more prominent affiliates may be one of the causes of persistent inequality in economic and social life. To the extent that prominent affiliates transfer status and resources, a pattern of cumulative advantage is set in motion for those with such relationships, leaving others behind. In this paper I propose that an opportunity to build relationships with more prominent affiliates can improve performance, but that the intellectual and social capital of a focal actor moderate the extent to which they can leverage such an opportunity to their advantage. Given that not everyone benefits equally, I further propose that not only do relationships with prominent affiliates create a wedge in between actual quality and outcomes, but they can also increase inequality amongst groups of less elite actors.

To test this proposition I make use of a unique natural experiment. The 2014 West African ebola epidemic provided scientists in countries most affected an unexpected opportunity to build relationships with more prominent affiliates. Scientists from around the world flooded into endemic countries looking for collaborators and research project assistance, and paid unique attention to the work and environment of scientists in these countries. I measure publication rates and international collaborations for endemic country scientists as compared with scientists in similar countries that weren't affected by the epidemic. The size of the effect is large. Endemic country scientists with relevant intellectual and social capital generate over 250% more journal impact factor weighted publications a year following the epidemic, and form an average of around fifteen more collaborations with scientists from OECD country institutions. Longer term, these positive benefits are maintained - and furthermore, endemic country scientists occupy different, and potentially more prominent, roles on scientific projects. However, those scientists in endemic countries without the relevant intellectual or social capital prior to the opportunity experienced no, or negative, effects of the epidemic. This implies increasing inequality within endemic country scientists as compared to non-endemic countries, and evidence for the concept brought into the forefront of our view of economic and social life many decades ago: 'the rich get richer and the poor get poorer'.

This manuscript makes several contributions. First, at a theoretical level, these results contribute to a central concept in sociology: the uncoupling between quality, status and performance. I provide evidence that the link between these concepts can be weakened by relationships with more prominent affiliates, and that heterogeneity in the ability to leverage such relationships can lead to self-reinforcing dynamics in the maintenance of a status distribution. At a more applied level the findings suggest that the standard approach to estimating the effect of relationships with more prominent affiliates on performance through measuring whether individuals with higher quality and quantity of such relationships perform better is likely to overstate the causal influence. This could be for two reasons: (1) it is difficult to control for underlying quality of an individual, and so it is hard to tell if a measurement of better performance is reflecting an underlying quality difference that leads to both an individual's ability to establish and maintain relationships or the impact of the relationship itself, and (2) causality could be flowing in the opposite direction: performance could determine relationships, which would bias any results based on observational evidence. For these reasons, much of the existing empirical literature could be overstating the true impact of relationships with prominent affiliates. This paper's finding that an opportunity can only be leveraged for advantage by those with relevant intellectual and social capital has important implications for our understanding of network based advantage and the design of social interventions. This is consistent with prior research that finds that peer effects are hard to replicate in field experiments due to post-randomization, endogenous sorting of relationships (Carrell et al 2013; Koning 2016; Hasan and Koning 2019).

Second, the results are consistent with theoretical and macro-level evidence suggesting that globalization can increase inequality amongst low-income countries. By using micro-level evidence at the level of individuals who are subject to an opportunity to build relationships with global counterparts, I can identify the role of both intellectual and social capital in determining the beneficiaries of globalization, and quantify resulting inequality that such an opportunity creates.

Third, to the best of my knowledge this is one of the first studies providing causal evidence of a channel by which scientific capacity in low-income countries can be developed. Despite the increasing attention from policy makers and donors on this important topic, we know surprisingly little about what works to build scientific capacity in low-income countries. I show that connections with more prominent scientists in higher income countries can help the careers of some scientists in countries with emerging ecosystems, but at the same time affect the system in unexpected ways.

The findings from this research have their limits. Pragmatically, the study is limited by the short time period available following the epidemic, and relies on publication outcomes, which may be a noisy reflection of true scientific capacity. The study also relies on a relatively small sample of 52 treated scientists. On a policy level, relationships between African and foreign collaborators can be problematic. These relationships are frequently described using such terms as 'parachute science', 'extractive relationships', and there are concerns of African collaborators assisting with field work and data collection, receiving little to none of the credit (Boshoff 2009). While this study is not able to tease apart involvement in a publication, or exploitative relationships, from actual scientific capacity, this is an important avenue for future research. Finally, this study is unable to discern the relative contribution of each proposed mechanism driving observed performance benefits: whether access to resources and knowledge, or endorsements and status transferral predominantly drives improvements in performance is an important question that deserves attention.

That said, I interpret the empirical results as providing support for the insights in the theoretical framework. Since Merton's seminal work (Merton 1968) on stratification in science, scholars have had an interest in the mechanisms driving such stratification in the scientific setting and others. While we know that those at the 'top' or the exclusive elite benefit from their position in the status ordering, we know surprisingly little about how those less elite can overcome the disadvantages of cumulative advantage. This study is a first step towards understanding the causal role of inter-status relationships in the trajectory of those less elite. I show that relationships with more prominent affiliates can improve performance, but that these benefits are reserved for those who already have relevant intellectual and social capital. This finding has significant implications for the design of policy measures aiming to promote development for groups outside of the elite, as well as for strategic measures implemented by those less elite. While cautious about making claims for the general applicability of these findings, particularly for settings without such high levels of uncertainty, I believe that the effects shown here can help provide a better understanding of why stratification remains pervasive across a variety of social systems, and the challenges that those less elite face in overcoming the disadvantages of status orderings.

References

Abadie, A, et al., 2017 *When Should You Adjust Standard Errors for Clustering?* NBER Working Paper No. w24003

Ahmadpoor, M., Jones, B.F., 2018 *Team Output and Individual Productivity in Science and Invention* Working Paper

Ahuja, G., 2000 *The Duality of Collaboration: Inducements and Opportunities in the Formation of Interfirm Linkages* Strategic Management Journal 21: 317-343

Allison, P.D., Long, J.S, Krauze, T.K., 1982 *Cumulative Advantage and Inequality in Science* American Sociological Review 47 (5): 615-625

Azoulay, P., Zivin, J.G., Wang, J., 2010 *Superstar Extinction* Quarterly Journal of Economics 125(2): 549-589

Azoulay, P., Stuart, T., Wang, Y., 2014 *Matthew: Effect or Fable?* Management Science 60(1): 92-109

Azoulay, P, Liu, C.C., Stuart, T.E., 2017 *Social Influence Given (Partially) Deliberate Matching: Career Imprints in the Creation of Academic Entrepreneurs* American Journal of Sociology 122(4): 1223-1271

Blau, P.M., 1964 *Exchange and Power in Social Life* New York: Wiley

Blau, P.M., 1977 *Inequality and Heterogeneity: A Primitive Theory of Social Structure* Vol. 7. New York: Free Press

Boshoff, N., 2009 *Neo-colonialism and Research Collaboration in Central Africa* Scientometrics 81: 413

Burt, R.S., 1995 *Structural Holes: The Social Structure of Competition* Harvard University Press

Burt, R.S. 2002 *Bridge Decay* Social Networks 24(4): 333-363

Burt, R.S., 2010 *Neighbor Networks: Competitive Advantage Local and Personal* Oxford University Press

Burton, M.D., Sorenson, J.B., Beckman, C.M 2002 *Coming from Good Stock: Career Histories and New Ventures Formation* In Social Structure and Organizations Revisited. Vol. 19, Research in the Sociology of Organizations, ed. Michael Lounsbury & Mark J. Ventresca, 229-262. Greenwich, CT: JAI Press

Carrell, S.E., Sacerdote, B.I., West, J.E., 2013 *From Natural Variation to Optimal Policy? The Importance of Endogenous Peer Group Formation* Econometrica 81(3): 855-882

Cole, J.R., Cole, S., 1973 *Social Stratification in Science* Chicago: University of Chicago Press

Coleman, J.S., 1988 *Social Capital in the Creation of Human Capital* American Journal of Sociology 94: S95-S120

Correll, SJ., et al. 2017 *It's the Conventional Thought That Counts: How Third-Order Inference Produces Status Advantage* American Sociological Review 82(2): 297-327

Ding, W.W., et al. 2010 *The Impact of Information Technology on Academic Scientists' Productivity and Collaboration Patterns* Management Science 56(9): 1439-1461

Donald, S.G., Lang, K., 2007 *Inference with Difference-in-Differences and Other Panel Data* The Review of Economics and Statistics 89(2): 221-233

Fitchett, J.R., et al., 2016 *Ebola Research Funding* Journal of Global Health 6: 020703

Goldsmith-Pinkham, P., Imbens, G.W., 2013 *Social Networks and the Identification of Peer Effects* Journal of Business & Economic Statistics 31(3): 253-264

Granovetter, M.S., 1973 *The Strength of Weak Ties* The American Journal of Sociology 78(6): 1360-1380

Goode, W.J., 1978 *The Celebration of Heroes: Prestige as a Social Control System* Berkeley, CA: University of California Press

Gould, R.V., 2002 *The Origins of Status Hierarchies: A Formal Theory and Empirical Test* American journal of sociology 107(5): 1143-1178

Gulati, R., 1999 *Network Location and Learning: The Influence of Network Resources and Firm Capabilities on Alliance Formation* Strategic Management Journal 20: 397-420

Hasan, S., Koning, R., 2019 *Prior Ties and the Limits of Peer Effects on Startup Team Performance* Strategic Management Journal 40(9): 1394-1416

Hara, N., et al. 2003 *An Emerging View of Scientific Collaboration: Scientists' Perspectives on Collaboration and Factors that Impact Collaboration* Journal of the American Society for Information science and Technology 54(10): 952-965

Hausman, J., Hall, B.H., and Griliches, Z., 1984 *Econometric Models for Count Data with an Application to the Patents-R&D Relationship* Econometrica 52(4): 909-938

Iacus, S.M., King, G., Porro, G., 2011 *Multivariate Matching Methods that are Monotonic Imbalance Bounding* Journal of the American Statistical Association 106(493): 345-361

Jackson, M.O., Wolinsky. A., 1996 *A Strategic Model of Social and Economic Networks* Journal of Economic Theory 71(1): 44-74

Jackson M.O., et al., 2008 *Social and Economic Networks* Volume 3 Princeton University Press, Princeton

Jones, B.F., 2014 *The Knowledge Trap: Human Capital and Development Reconsidered* Working Paper

Jones, B., 2009 *The Burden of Knowledge and the 'Death of the Renaissance Man': Is Innovation Getting Harder?* Review of Economic Studies 76(1): 283-317

Jonkers, K., Tjissen, R., 2008 *Chinese Researchers Returning Home: Impacts of International Mobility on Research Collaboration and Scientific Productivity* Scientometrics 77: 309 - 333

Kim, J.W., King, B.G., 2014 *Seeing Stars: Matthew Effects and Status Bias in Major League Baseball Umpiring* Management Science 60(11): 2619-2644

Koning, R., 2016 *Do Network Dynamics Undermine Idea-based Network Advantages? Experimental Results from an Entrepreneurship Bootcamp* Working Paper

Kremer, M., 2006 *Globalization of Labor Markets and Inequality* Brookings Trade Forum: 211-228

Latour, B., 1987 *Science in Action* Cambridge MA: Harvard University Press

Levin, S.G., Stephan, P.E., 1991 *Research Productivity over the Life Cycle: Evidence for Academic Scientists* American Economic Review 81: 114-132

Levinthal, D.A., 1997 *Adaptation on Rugged Landscapes* Management Science 43(7): 934-950

Long, J.S., Allison, P.D., McGinnis, R., 1979 *Entrance into the Academic Career* American Sociological Review 44(5): 816-830

Manski, C.F., 1993 *Identification of Endogenous Social Effects: The Reflection Problem* The Review of Economic Studies 60(3): 531-542

Marsden, P.V., 1983 *Restricted Access in Networks and Models of Power* American Journal of Sociology 88: 686-717

Maskin, E., 2015 *Why Haven't Global Markets Reduced Inequality in Emerging Economies?* The World Bank Economic Review 29(1): S48-S52

Merton, R.K., 1968 *The Matthew Effect in Science* Science 159(3810): 56-63

Merton, R.K., 1972 *Insiders and Outsiders: A Chapter in the Sociology of Knowledge* American Journal of Sociology 78(1): 9-47

Moran, M., et al 2014 *Neglected Disease Research and Development: Emerging Trends* G-Finder 2014 Policy Cures

Mouw T., 2006 *Estimating the Causal Effect of Social Capital: A Review of Recent Research* Annual Review of Sociology 32: 79 -102

Mutters, N.T., et al., 2018 *Evaluation of the Scientific Impact of the Ebola Epidemic: A Systematic Review* Clinical Microbiology and Infection 24: 573-576

Nelson, R., Winter, S., 1982 *An Evolutionary Theory of Technical Change* Cambridge, Ma, Beknap Harvard

Pfeffer, J., 1983 *Organizational Demography* Research in Organizational Behavior

Podolny, J.M., 1993 *A Status-Based Model of Market Competition* American journal of sociology 98(4): 829-872

Reagans, R.E., Burt, R.S., 1998 *Homophily, Legitimacy, and Competition: Bias in Manager Peer Evaluations* Annual meetings of the American Sociological Association

Reagans, R., B. McEvily. 2003 *Network Structure and Knowledge Transfer: The Effects of Cohesion and Range* Admin. Sci. Quart.48240–267

Rivera, M.T., Soderstrom, S.B., Uzzi, B., 2010 *Dynamics of Dyads in Social Networks: Assortative, Relational, and Proximity Mechanisms* Annual Review of Sociology 36: 91-115

Simcoe, T.S., Waguespack, D.M, 20111 *Status, Quality and Attention: What's in a (Missing) Name?* Management Science 57(2): 274-290

Spence, A.M., 1974 *Market Signaling* Cambridge, MA: Harvard University Press

Sorenson, O., Waguespack, D.M., 2006 *Social Structure and Exchange: Self-confirming Dynamics in Hollywood* Administrative Science Quarterly 51(4): 560-589

Stuart, T.E., Hoang, H., Hybels, R.C., 1999 *Interorganizational Endorsements and the Performance of Entrepreneurial Ventures* Administrative Science Quarterly 44(2): 315-349

Stuart, T.E., Sorenson, O., 2007 *Strategic Networks and Entrepreneurial Ventures* Strategic Entrepreneurship Journal 1: 211-227

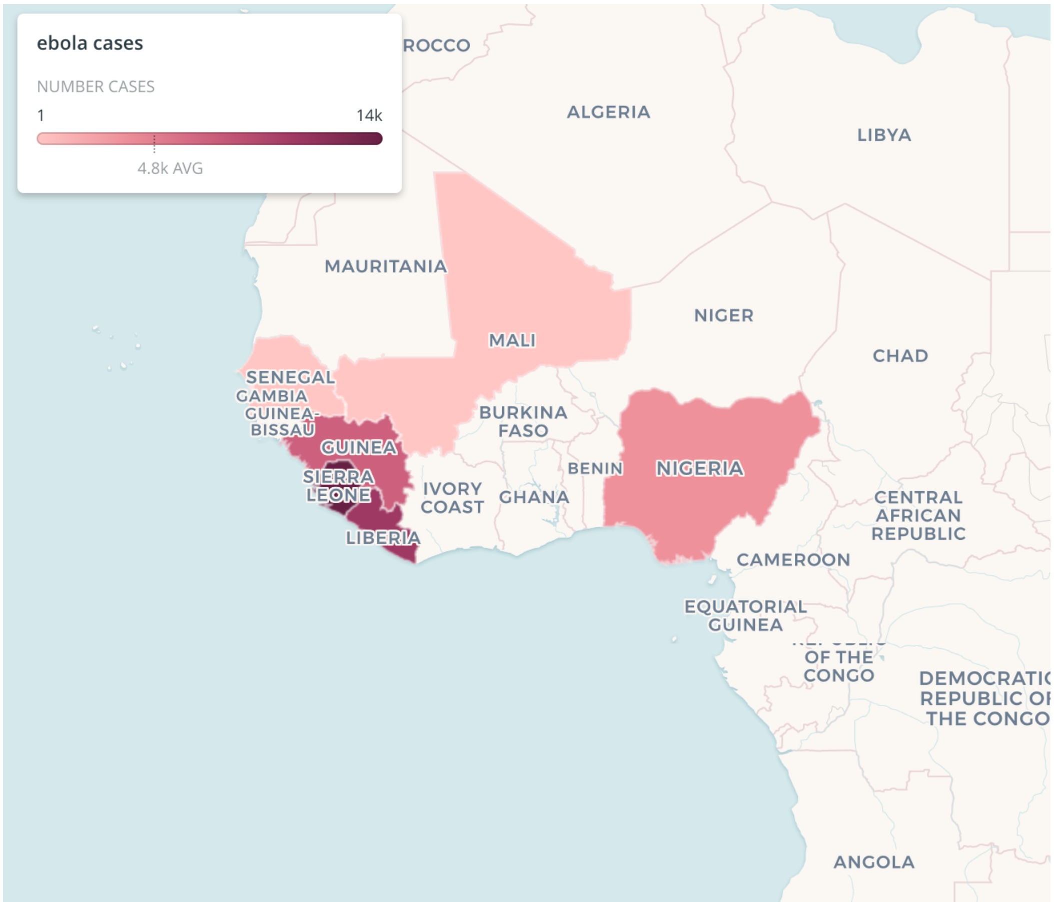
Wagner, C.S., Leydesdorff, L., 2005 *Network Structure, Self-organization, and the Growth of International Collaboration in Science* Research Policy 34(10): 1608-1618

Van Raan, A.F.J., 1998 *The Influence of International Collaboration of the Impact of the Research Results* Scientometrics 42(3): 423-428

Zuckerman, H., 1970 *Stratification in American Science* pp 235-57 in E.O. Laumann (ed.). Social Stratification. Indianapolis: Bobbs-Merrill

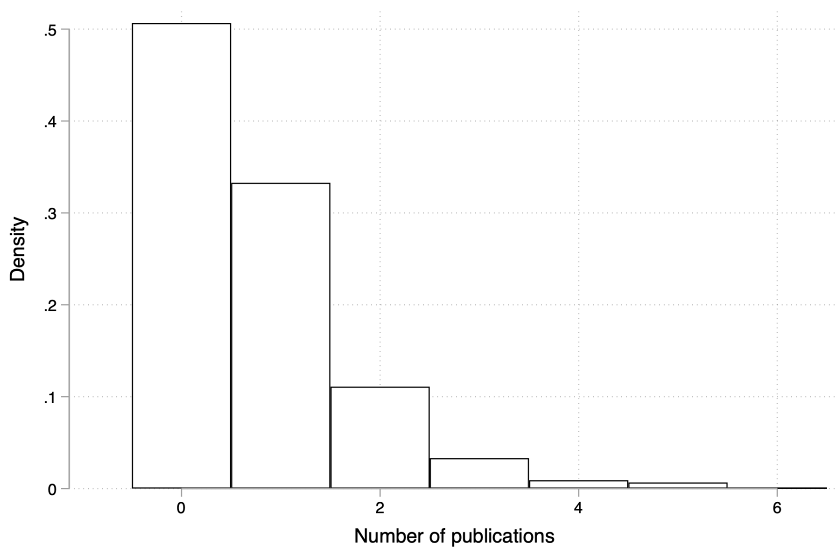
Zuckerman, H., 1988 *The Sociology of Science* In Handbook of Sociology, edited by N.J. Smelser. Newbury Park, CA: Sage

Figure 1: 2014 West Africa ebola outbreak - cases by country



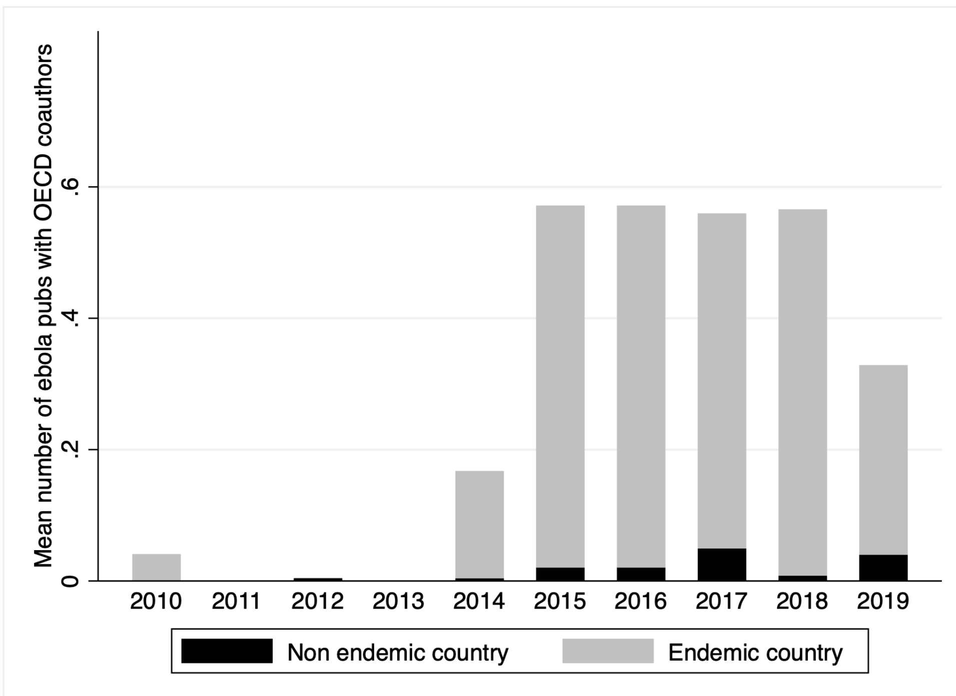
Note: This figure represents official World Health Organization statistics of number of total cases (suspected, probable, confirmed) by the end of the outbreak in 2016.

Figure 2: Histogram of Annual Publication Rate



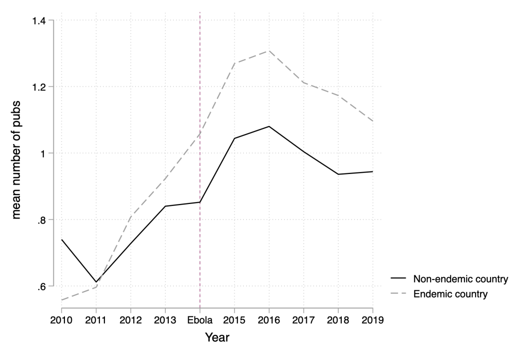
Note: I compute the number of publications per year authored by the 302 sample treated and control scientists prior to the epidemic (2010-2013).

Figure 3: Mean Number Ebola Publications with OECD Coauthors Authored by Sample Scientists

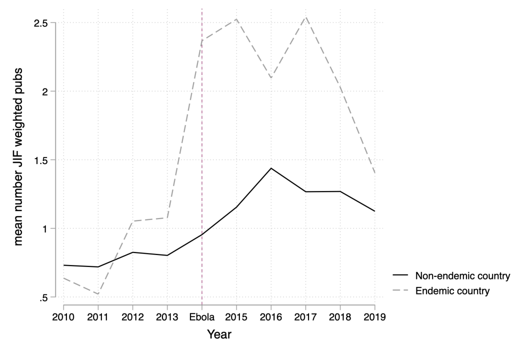


Note: Average number of ebola publications co-authored between OECD and sample scientists in the year of observation are calculated for endemic country and control country scientists and plotted above. The lighter gray bars correspond to the mean number of ebola publications for endemic country scientists, and the darker black bars correspond to the mean number of ebola publications for control country scientists. The ebola epidemic struck in 2014.

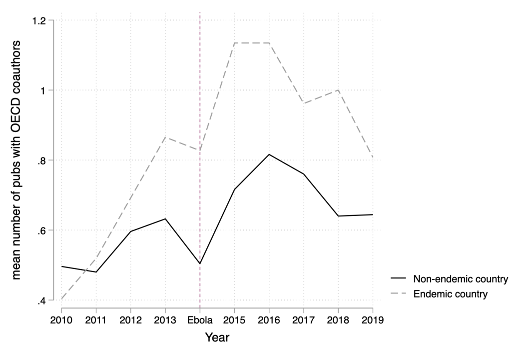
Figure 4: Outcomes for Endemic Country Scientists vs Non Endemic, Control Country Scientists Following Ebola Epidemic



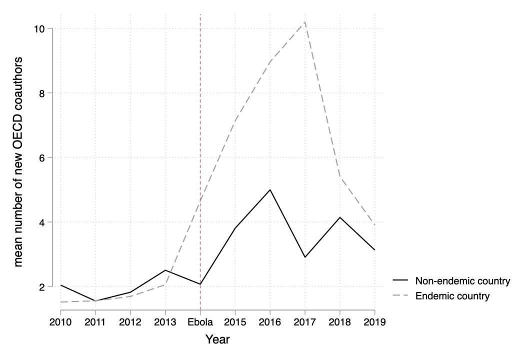
(a) Number of publications



(b) Number of Journal Impact Factor (JIF) weighted publications



(c) Publications with OECD coauthors



(d) Number new OECD coauthors

Note: Raw averages of outcomes each year are calculated for endemic country scientists and control country sample scientists and plotted above. The solid lines correspond to mean outcomes for endemic country scientists, and the dashed correspond to mean outcomes for control country scientists. The vertical dotted line illustrates the year that the ebola epidemic struck (2014).

Table 1: Details of West and Central African Countries Included in Analysis

Country	Nb. ebola cases (confirmed and suspected) (2014-2016)	GDP per capita	Population (in millions)	Nb. biomedical/social scientists in 2014
Treated				
Sierra Leone	14,124	788	6.2	28
Liberia	10,678	461	4.4	6
Guinea	3,814	550	12.0	27
Control				
Nigeria	20	3,185	178.5	3,327
Mali	8	766	15.8	133
Senegal	1	1,071	14.6	434
Gabon	0	10,067	1.7	88
Angola	0	5,936	22.1	12
Congo, Republic	0	3,100	4.6	86
Cote d'Ivoire	0	1,646	20.8	359
Ghana	0	1,462	26.4	726
Cameroon	0	1,426	22.8	455
Mauritania	0	1,270	3.9	7
Chad	0	1,053	13.2	7
Benin	0	825	10.6	397
Burkina Faso	0	720	17.4	306
Togo	0	646	6.9	105
Guinea-Bissau	0	586	1.7	12
Congo, Democratic Republic	0	475	69.4	89
Niger	0	441	18.5	53
Gambia, The	0	423	1.9	78
Central African Republic	0	379	4.7	23

Note: Details of the countries included in the sample before the matching procedure are given in the table. The column 'nb. biomedical or social scientists in country at time of ebola outbreak' provides numbers of all possible treated and control scientists in each country on which the matching procedure will take place. This full set of possible study scientists is identified as those publishing in the Elsevier Scopus database prior to 2014 and publishing at least once after 2012 (to exclude retired scientists). Further inclusion criteria is that scientists publish at least three times during their entire publication history and are first or last author at least once on a publication (to exclude lab technicians). Their country of residence is determined as a rule of over 75% of their affiliations being based in a particular country between 2010 and 2014.

Table 2: Study Sample

Country	Nb. scientists in study sample
Treated	
Sierra Leone	25
Guinea	21
Liberia	6
Total	52
Control	
Mali	62
Congo, Democratic Republic	53
Togo	50
Gambia, The	38
Niger	33
Central African Republic	9
Guinea-Bissau	5
Total	250

Note: This final study sample is a subset of the sample provided in Table 1 following the coarsened exact matching procedure. The matching procedure identifies comparable treated and control scientists based on covariates such as country specific features, and researcher specific features such as career age and publication record, and excludes scientists for whom an appropriate match cannot be found.

Table 3: Summary Statistics for Study Scientists the Year Prior to the Ebola Epidemic

	Control Scientists (N = 250)					Treated Scientists (N = 52)				
	mean	median	std. dev.	min.	max.	mean	median	std. dev.	min.	max.
Career age	9.56	7	5.85	4	29	11.90	10	7.14	4	29
Any publication	0.59	1	0.49	0	1	0.63	1	0.49	0	1
Nb. publications	0.84	1	0.91	0	5	0.92	1	1.04	0	5
Nb. JIF weighted publications	0.80	0.41	1.05	1	5.62	1.08	0.53	1.47	0	6.09
Any last author publication	0.088	0	0.28	0	1	0.077	0	0.27	0	1
Nb. last author publications scientist	0.10	0	0.37	0	3	0.077	0	0.27	0	1
Any publication with OECD coauthors	0.48	0	0.50	0	1	0.58	1	0.50	1	1
Nb. publications with OECD coauthors	0.63	0	0.79	0	1	0.87	1	1.07	0	5
Nb. new OECD coauthors	2.50	0	8.18	0	111	2.06	0	3.08	0	11
Any publication in tropical diseases	0.24	0	0.43	0	1	0.29	0	0.46	0	1
Nb. publications in tropical diseases	0.32	0	0.67	0	4	0.35	0	0.68	0	4
Any publication in viral hemorrhagic fever (2010-2013)	0.008	0	0.089	0	1	0.077	0	0.27	0	1

Note: The sample consists of 302 Central and West African scientists who were actively publishing at the time of the ebola outbreak (2014). All statistics are measured using scientist year level data gathered from the Elsevier Scopus database, and measurements are made the year prior to the epidemic (at the end of 2013). Nb. of publications are publications in 2013; any publications represents a dummy variable for whether the sample scientist has a publication in 2013.

Table 4: Impact of Ebola Epidemic on Endemic Country Scientists' Publication Output

Dependent variable	(1) Nb. pubs	(2) Nb. JIF weighted pubs	(3a) Nb. pubs with OECD	(3b) Nb. pubs without OECD	(4) Nb. new OECD coauthors
AFTER EPIDEMIC × ENDEMIC COUNTRY	0.20	0.52**	0.24**	0.12	0.79***
	(0.17)	(0.25)	(0.12)	(0.28)	(0.21)
Mean of dependent variable	0.90	1.13	0.66	0.24	3.21
Log likelihood	-2856	-3605	-2276	-1014	-9326
Nb. of scientists (nb. of treated scientists)	302 (52)	301 (52)	287 (49)	188 (33)	285 (48)
Nb. of scientists × year observations	3,020	3,010	2,870	1,880	2,850

[a] Estimates stem from quasi-maximum likelihood fixed effects Poisson specifications with dependent variables being counts of outcomes per scientist in the year of observation. All models incorporate a full suite of calendar year and scientist fixed effects and six career age category indicator variables. Exponentiating the coefficients and differencing from one yields numbers interpretable as elasticities. The number of observations varies slightly across columns because the conditional fixed effects specification drops observations corresponding to scientists for which there is no variation in activity over the entire observation period.
[b] Heteroskedastic robust standard errors, clustered at the individual scientist level, are given in parentheses. *, **, *** denote statistical significance at p-values of 0.1, 0.05 and 0.01 respectively

Table 5: Breakdown of Impact of Ebola Epidemic on Endemic Country Scientists' Publication Output by Pre-Epidemic Intellectual Capital

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dependent variable	Nb. pubs		Nb. JIF weighted pubs		Nb. pubs with OECD		Nb. new OECD coauthors	
Scientist subsample	Tropical Disease Record	No Tropical Disease Record	Tropical Disease Record	No Tropical Disease Record	Tropical Disease Record	No Tropical Disease Record	Tropical Disease Record	No Tropical Disease Record
AFTER EPIDEMIC × ENDEMIC COUNTRY	0.50*	-0.37	1.00***	-0.62**	0.60**	-0.49	1.11***	-0.27**
	(0.21)	(0.31)	(0.28)	(0.29)	(0.21)	(0.36)	(0.34)	(0.11)
Mean of dependent variable	1.06	0.76	1.55	0.78	0.86	0.50	4.62	2.02
Log likelihood	-1412	-1424	-2044	-1511	-1223	-1030	-5457	-3660
Chi^2		5.49, p<.05		16.22, p<.05		6.89, p<.05		6.13, p<.05
Nb. of scientists (nb. of treated scientists)	138 (26)	164 (26)	138 (26)	163 (26)	136 (26)	151 (23)	135 (25)	150 (23)
Nb. of scientists × year observations	1,380	1,640	1,380	1,630	1,360	1,510	1,350	1,500

[a] The sample is divided into two sub-samples: those who have authored any publications in the four years before the epidemic in neglected tropical disease subject areas in columns 1,3,5,7, and the rest of the sample in columns 2,4,6,8. The same specification is run on each sample. The Chi^2 value represents significance tests between the two models that use distinct samples for a given outcome. Pairs with p-values less than 0.05 signifies that the relationship between the dependent variable and the independent variable (after epidemic × endemic country) differs by subset of the sample at the 0.05 significance level.
[b] Estimates stem from quasi-maximum likelihood fixed effects Poisson specifications with dependent variables being counts of outcomes per scientist in the year of observation. All models incorporate a full suite of calendar year and scientist fixed effects and six career age category indicator variables. Exponentiating the coefficients and subtracting one yields numbers interpretable as elasticities. The number of observations varies slightly across columns because the conditional fixed effects specification drops observations corresponding to scientists for which there is no variation in activity over the entire observation period.
[c] Heteroskedastic robust standard errors, clustered at the individual scientist level, are given in parentheses. *, **, *** denote statistical significance at p-values of 0.1, 0.05 and 0.01 respectively

Table 6: Breakdown of Impact of Ebola Epidemic on Endemic Country Scientists' Publication Output by Pre-Epidemic Social Capital

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dependent variable	Nb. pubs		Nb. JIF weighted pubs		Nb. pubs with OECD		Nb. new OECD coauthors	
Scientist subsample	OECD collab. record	no sig. OECD collab. record	OECD collab. record	no sig. OECD collab. record	OECD collab. record	no sig. OECD collab. record	OECD collab. record	no sig. OECD collab. record
AFTER EPIDEMIC × ENDEMIC COUNTRY	0.29	0.10	0.58*	0.47	0.33	0.094	1.09***	0.26
	(0.23)	(0.24)	(0.32)	(0.40)	(0.24)	(0.28)	(0.31)	(0.60)
Mean of dependent variable	0.81	1.06	1.15	1.09	0.71	0.59	3.35	2.97
Log likelihood	-1706	-1139	-2373	-1193	-1552	-706	-6249	-2821
Chi^2		0.31, p>.05		0.05, p>.05		0.41, p>.05		1.51, p>.05
Nb. of scientists (nb. of treated scientists)	191 (36)	111 (16)	190 (36)	111 (16)	191 (36)	96 (13)	189 (35)	96 (13)
Nb. of scientists × year observations	1,910	1,110	1,900	1,110	1,910	960	1,890	960

[a] The sample is divided into two sub-samples: those who have published with OECD based collaborators in all of their publications in the four years prior to the epidemic in columns 1,3,5,7, and the rest of the sample in columns 2,4,6,8. The same specification is run on each sample. The Chi^2 value represents significance tests between the two models that use distinct samples for a given outcome. Pairs with p-values less than 0.05 signifies that the relationship between the dependent variable and the independent variable (after epidemic × endemic country) differs by subset of the sample at the 0.05 significance level.
[b] Estimates stem from quasi-maximum likelihood fixed effects Poisson specifications with dependent variables being counts of outcomes per scientist in the year of observation. All models incorporate a full suite of calendar year and scientist fixed effects and six career age category indicator variables. Exponentiating the coefficients and subtracting one yields numbers interpretable as elasticities. The number of observations varies slightly across columns because the conditional fixed effects specification drops observations corresponding to scientists for which there is no variation in activity over the entire observation period.
[c] Heteroskedastic robust standard errors, clustered at the individual scientist level, are given in parentheses. *, **, *** denote statistical significance at p-values of 0.1, 0.05 and 0.01 respectively

Table 7: Breakdown of Impact of Ebola Epidemic on Endemic Country Scientists' Publication Output by Pre-Epidemic Intellectual and Social Capital

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dependent variable	Nb. JIF weighted pubs				Nb. new OECD coauthors			
Scientist subsample	Tropical Disease and OECD Coauthor Record	Tropical Disease but no OECD Coauthor Record	OECD Coauthor but no Tropical Disease Record	No Tropical or OECD Coauthor Record	Tropical Disease and OECD Coauthor Record	Tropical Disease but no OECD Coauthor Record	OECD Coauthor but no Tropical Disease Record	No Tropical or OECD Coauthor Record
AFTER EPIDEMIC × ENDEMIC COUNTRY	1.30***	0.70	-0.52	-0.96**	1.50***	0.48	0.088	-2.57**
	(0.45)	(0.19)	(0.33)	(0.38)	(0.39)	(0.67)	(0.45)	(1.01)
Mean of dependent variable	1.51	1.64	0.84	0.67	4.53	4.79	2.32	1.53
Log likelihood	-1299	-702	-1020	-463	-3505	-1767	-2589	-845
Chi^2		1.05, p>.05	13.25 p<.05	17.64, p<.05		1.74, p>.05	5.67, p<.05	12.42, p<.05
Nb. of scientists (nb. of treated scientists)	89 (16)	49 (10)	101 (20)	62 (6)	88 (15)	47 (10)	101 (20)	49 (3)
Nb. of scientists × year observations	890	490	1,010	620	880	470	1,010	490

[a] The sample is divided into four exclusive sub-samples: (i) scientists with any neglected disease experience and an OECD coauthor record prior to the epidemic in columns 1 and 5; (ii) scientists with a record in neglected tropical disease research but without always having coauthored with OECD based scientists in columns 2 and 6: (iii) scientists who have coauthored all of their publications with OECD based scientists, but with no record in neglected tropical disease research in columns 3 and 7; and (iv) scientists with neither neglected disease experience nor having coauthored all of their publications with OECD based scientists in columns 4 and 8.
[b] Estimates stem from quasi-maximum likelihood fixed effects Poisson specifications with dependent variables being counts of outcomes per scientist in the year of observation. All models incorporate a full suite of calendar year and scientist fixed effects and six career age category indicator variables. Exponentiating the coefficients and subtracting one yields numbers interpretable as elasticities. The number of observations varies slightly across columns because the conditional fixed effects specification drops observations corresponding to scientists for which there is no variation in activity over the entire observation period.
[c] Heteroskedastic robust standard errors, clustered at the individual scientist level, are given in parentheses. *, **, *** denote statistical significance at p-values of 0.1, 0.05 and 0.01 respectively

Table 8: Persistent Impact of Ebola Epidemic on Endemic Country Scientists' Publication Output

Dependent variable	(1) Any pub	(2a) First author pub	(2b) Middle author pub	(2c) Last author pub	(3) Hit pub
ENDEMIC COUNTRY	0.20	0.039	0.0094	0.87*	0.81**
	(0.32)	(0.29)	(0.31)	(0.47)	(0.39)
Mean of dependent variable	0.57	0.47	0.54	0.054	0.11
Log likelihood	-425	-134	-430	-129	-219
Nb. of scientists (nb. of treated scientists)	161 (30)	161 (30)	161 (30)	161 (30)	161 (30)
Nb. of scientists $\times$ year observations	644	644	644	644	644

[a] Using a reduced sample of those who published in 2014-2015 with OECD coauthors, estimates stem from logit specifications with dependent variables being dummy of outcomes in a given year per scientist post 2016. All models incorporate controls for age bracket, year fixed effects and full suite of scientist level covariates (same as used for matching procedure) on productivity, research area, and international collaborations. Coefficients can be interpreted as log odds. Exponentiating the coefficient gives the odds ratio. So - for model (2c) the likelihood of being a last author on a publication is $\exp(0.87) = 2.4$ times more for those in endemic countries than the control countries.
[b] Heteroskedastic robust standard errors, clustered at the individual scientist level, are given in parentheses. *, **, *** denote statistical significance at p-values of 0.1, 0.05 and 0.01 respectively

Table 9: Sensitivity Checks

Dependent variable = Nb. JIF weighted pubs	(1) Benchmark specification	(2) Placebo test (only pre-period)	(3) Without most productive scientists
AFTER EPIDEMIC × ENDEMIC COUNTRY	0.52***	0.25	0.64***
	(0.11)	(0.25)	(0.31)
Mean of dependent variable	1.13	0.78	0.99
Log likelihood	-3605	-804	-2918
Nb. of scientists (nb. of treated scientists)	301 (52)	292 (52)	263 (43)
Nb. of scientists × year observations	3,010	1,168	2,630

[a] Estimates stem from quasi-maximum likelihood fixed effects Poisson specifications with dependent variables being counts of journal impact factor (JIF) weighted publications per scientist per year. Column (1) provides the baseline estimation as previous regressions with the full sample. Column (2) takes just the pre-epidemic period of data and moves the treatment date to 2012 – with endemic country scientists treated. Column (3) removes scientists from treated and control who are in the 90th percentile of publication output prior to the epidemic. All models incorporate a full suite of calendar year and scientist fixed effects and six career age category indicator variables. Exponentiating the coefficients and subtracting one yields numbers interpretable as elasticities. The number of observations varies slightly across columns because the conditional fixed effects specification drops observations corresponding to scientists for which there is no variation in activity over the entire observation period.

[b] Heteroskedastic robust standard errors, clustered at the individual scientist level, are given in parentheses. *, **, *** denote statistical significance at p-values of 0.1, 0.05 and 0.01 respectively

Appendix A: Construction of the Control Group

I detail the procedure implemented to identify the control scientists that help to account for life-cycle and secular trends in the difference-in-differences specification. Publication outcomes might be subject to life-cycle patterns, with outcomes reflecting the trends of the age of the scientist. Also - scientific productivity, particularly in Africa, is rapidly changing over time. Therefore it is important to fully capture these time-varying omitted variables.

To address this concern, I create a sample of control scientists to account for time varying variables in the difference-in-differences specification. Specifically I identify control scientist(s) who is(are) 'similar' to each treated scientist. The control scientists are selected from a universe of possible biomedical or social scientists who are based West or Central African countries that are not considered highly epidemic during the 2014 West Africa ebola epidemic.

The universe of possible control scientists is generated using affiliation data from Elsevier Scopus publication database with inclusion criteria such that the scientist must have published at least three times in their lifetime and at least once as first or last author (to remove technicians or incidental publishers) and published at least once in biomedical or social sciences. The country of each scientist is determined as being the country in which they are affiliated with in a given time period in over 75% of their publications (to avoid visiting or honorary appointments).

The list of covariates used to identify 'similar' control scientists for each treated scientist such that the following conditions are met:

1. treated scientists are located in similar countries in terms of GDP and size of the scientific workforce at the time of the epidemic;

2. treated scientists exhibit no differential output trends relative to control collaborators up to the time of the epidemic;

3. treated scientists exhibit no differential trends in terms of international, particularly OECD, collaborations relative to control collaborators up to the time of the epidemic;

4. treated scientists exhibit no differential trends in terms of their field of study relative to control collaborators up to the time of the epidemic;

5. the distribution of career age at the time of the epidemic are for similar treated and control scientists.

Coarsened exact matching. To meet these goals, I implement the nonparametric 'coarsened exact matching' (CEM) procedure (Iacus et al 2011) to identify at least one control scientist for each treated scientist. The first step is to select a set of covariates on which to guarantee balance, and the second is to create a large number of (coarse) strata that covers the entire support of the joint distribution of the covariates in the previous step. In a third step, each observation is allocated to a stratum and for each treated observation, control observations are selected from the same stratum. If the treated observation is unmatched it is removed from the sample.

Implementation I identify controls based on the following set of covariates: GDP per capital and number of scientists at the end of 2013; career age in 2014, the number of publications in the years 2010-2013 inclusive, the number of journal impact factor weighted publications in each year 2010-2013 inclusive, a dummy variable if the scientist published with OECD based collaborators in the years 2010-2013 inclusive, and the percentage of publications in neglected tropical diseases in years 2010-2013 inclusive.

Appendix B: Alternative Functional Form Specifications

Table B1: Alternative Functional Form Specifications

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dependent variable	IHS transformed JIF weighted pubs				log(Y+1) transformed JIF weighted pubs			
Scientist subsample	Tropical Disease and OECD Coauthor Record	Tropical Disease but no OECD Coauthor Record	OECD Coauthor but no Tropical Disease Record	No Tropical or OECD Coauthor Record	Tropical Disease and OECD Coauthor Record	Tropical Disease but no OECD Coauthor Record	OECD Coauthor but no Tropical Disease Record	No Tropical or OECD Coauthor Record
AFTER EPIDEMIC × ENDEMIC COUNTRY	0.55**	0.45*	-0.19	-0.28**	0.45**	0.36*	-0.15	-0.22**
	(0.24)	(0.25)	(0.13)	(0.12)	(0.19)	(0.20)	(0.10)	(0.094)
Mean of dependent variable	1.51	1.64	0.84	0.67	1.51	1.64	0.84	0.67
R^2 (adjusted)	0.23	0.31	0.20	0.26	0.23	0.30	0.20	0.27
Chi^2		1.10, p>.05	7.75, p<.05	10.07, p<.05		0.12, p>.05	7.89, p<.05	10.24, p<.05
Nb. of scientists (nb. of treated scientists)	89 (16)	49 (10)	102 (20)	62 (6)	89 (16)	49 (10)	102 (20)	62 (6)
Nb. of scientists × year observations	890	470	1,020	490	890	490	1,020	620

[a] The sample is divided into four exclusive sub-samples: (i) scientists with neglected disease experience and an OECD coauthor record prior to the epidemic in columns 1 and 5; (ii) scientists with a record in neglected tropical disease research but without always having coauthored with OECD based scientists in columns 2 and 6: (iii) scientists who have coauthored all of their publications with OECD based scientists, but with no record in neglected tropical disease research in columns 3 and 7; and (iv) scientists with neither neglected disease experience nor having coauthored all of their publications with OECD based scientists in columns 4 and 8. [b] Estimates stem from fixed effects Ordinary Least Squares specifications with dependent variables being inverse hyperbolic sine transformation of journal impact factor (JIF) weighted sum of publications in columns 1-4, and log(Y+1) transformation of JIF weighted publications (Y) in columns 5-8.
[c] All models incorporate a full suite of calendar year and scientist fixed effects and six career age category indicator variables. Exponentiating the coefficients and subtracting one yields numbers interpretable as elasticities. The number of observations varies slightly across columns because the conditional fixed effects specification drops observations corresponding to scientists for which there is no variation in activity over the entire observation period.
[d] Heteroskedastic robust standard errors, clustered at the individual scientist level, are given in parentheses. *, **, *** denote statistical significance at p-values of 0.1, 0.05 and 0.01 respectively

Appendix C: Additional Outcomes

Additional analysis is carried out on further outcomes, namely - author position and topic of publications. First, the probability that a scientist is first, one of the middle, or last author is measured as a proxy for their contribution to the article (with first authors generally contributing the ground work and last authors generally contributing intellectual oversight as well as funding). Because of the rare occurrences (particularly of last authors) in the study sample, I use a dummy variable that takes the value of 1 if a scientist is first, any middle, or last author on a publication in a given year, and 0 otherwise. Finally, I carry out key word searches of the title, abstract and keywords of each publication for a sample scientist in an observation year to ascertain their research area. Specifically I am interested in whether scientists are switching their research areas in response to the epidemic. I measure the probability that a scientist publishes in neglected tropical diseases (the broader category of diseases that ebola falls under), and non-neglected tropical disease areas. Table 11 provides the results of a linear probability model estimating the change in the likelihood of these outcomes after the epidemic, as compared to control scientists.

Table C1: Impact of Ebola Epidemic on Endemic Country Scientists' Publication Outcomes - Additional Outcome Measures

	(1) Any pub	(2a) First author pub	(2b) Middle author pub	(2c) Last author pub	(3a) Tropical disease pub	(3b) Non tropical disease pub
AFTER EPIDEMIC × ENDEMIC COUNTRY	0.017	0.0066	-0.011	0.025	0.082*	0.17
	(0.051)	(0.049)	(0.054)	(0.026)	(0.044)	(0.051)
Mean of dependent variable	0.50	0.55	0.44	0.050	0.25	0.32
R^2	0.11	0.14	0.13	0.10	0.31	0.20
Nb. of scientists (nb. of treated scientists)	302 (52)	302 (52)	302 (52)	302 (52)	302 (52)	302 (52)
Nb. of scientists × year observations	3,020	3,020	3,020	3,020	3,020	3,020

[a] Estimates stem from fixed effects linear probability model specifications with dummy outcomes. All models incorporate a full suite of calendar year and scientist fixed effects and six career age category indicator variables. The coefficients can be interpreted as changes in probabilities in the outcome occurring.
[b] Heteroskedastic robust standard errors, clustered at the country level, are given in parentheses. *, **, *** denote statistical significance at p-values of 0.1, 0.05 and 0.01 respectively

Chapter 3

3 Building Bridges: The Impact of Return Migration by African Scientists

Abstract

Despite significant interest in the potential for 'returnee' scientists moving back to developing countries to connect developed and developing countries, prior work has found limited evidence of success. I shift the focus to the broader network of the returnee, and study the extent to which the return home of American-trained HIV researchers to African institutions impacts publication outcomes of non-migrant scientists in Africa. I find that following the arrival of a returnee in their institution, non-migrants experience increased productivity, mostly in HIV research. I find strong evidence that the mechanism driving this effect is that of the returnee providing a bridge to their central connections and subsequent knowledge and resources thus affecting outcomes. In settings where 'outsiders' struggle to access knowledge and resources that are usually reserved for exclusive 'insiders', this kind of bridge in the network can help through providing legitimacy to the outsiders. These findings inform a network perspective on the consequences of the mobility of skilled individuals, the development of national innovation ecosystems, and the globalization of knowledge production.

> 'I was a big go-between...if people (at my institution) are specialized in different ways from me, and people from the United States reach out to me, I introduce them.'
>
> *Returning African scientist*

3.1 Introduction

The consequences of the global movement of high-skilled workers has animated much research.[7] In particular, return migration of high-skilled workers from more developed countries to developing countries has received renewed interest in recent years (Borjas and Bratsberg 1996; Zucker and Darby 2007; Dustmann et al 2011; Gaule 2014). Great hope is attached to these return migrants and their role in transforming their home country economy through brokering access to knowledge and resources in more developed countries (Saxenian 2006).

Celebrated cases of returnees contributing towards home country economies (Saxenian 2006), and empirical evidence of returnees bringing back knowledge and resources (Jonkers and Tjissen 2008; Choudhury 2015; Gianetti et al 2015) support the image of return migrants as successful brokers. But another line of research provides evidence that returnees face challenges in their brokerage role due to a variety of individual and interpersonal barriers, including the presence of other returnees, home country xenophobia, and challenges in maintaining ties at home and abroad (Obukhova 2012; Wang 2015). Returnees working in knowledge production may face additional challenges due to the limited availability of resources and collaborators at home, geographic concentration of knowledge flows, and bias based on institutional or geographic affiliation. Indeed, evidence of declining productivity of scientists as they move home to developing countries seems to confirm that benefits to brokerage are limited in this setting (Gibson and McKenzie 2014; Kahn and MacGarvie 2016a).

It may be, however, that this reflects a narrow view of brokerage and diverts our attention from a broader set of causal pathways by which bridges across disparate parts of a system can benefit those who are impacted by the bridge. In particular, and following Burt's notion of 'second-hand brokerage,' it could be that while the returnees themselves benefit relatively little from the connections they make, their associates in developing countries do indeed benefit. Evidence that a broker's connections, and subsequent access to knowledge and resources, can be shared with their associates is limited (Burt 2007,

[7] Kerr 2008; Oettl and Agrawal 2008; Hunt and Gauthier-Loiselle 2010; Agrawal et al 2011; Borjas and Doran 2012; Kogut and Macpherson 2011; Moser et al 2014; Ganguli 2015

2010). However, there is strong theoretical and empirical reason to expect such a benefit when the bridge allows 'outsiders' in a system access to connections, knowledge or resources otherwise restricted to legitimate 'insiders' (Burt 1997, 1998, 2010; Stuart et al., 1999). Moreover, recent research (Fry 2019) documents such sharing in the context of scientific collaboration across the divide between developed and developing countries.

By developing and applying Burt's idea of second-hand brokerage in the context of global science, which exhibits a core/periphery network structure (Crane 1965; Cole and Cole 1973; Zuckerman 1988; Zelnio 2012), we can more clearly illuminate how brokers impact systems more generally. Core/periphery structures are characterized by densely connected core actors (insiders) and loosely connected peripheral actors (outsiders) (Borgatti and Everett 1999). The selective core represents influential actors and their position is associated with privilege, control and prestige (Clauset et al 2015). An actor's network position relative to the core, and thus their access to central connections, resources and knowledge, has consequences for creative and innovative output (Cattani et al 2014; Cattani and Ferriani 2008). An examination of the extent to which periphery actors can access connections, knowledge and resources of an actor who forms a bridge between the core and the periphery - a "core/periphery bridge" - holds promise in this setting and others.

It is difficult to identify the causal impact of such sharing. Actors in a network may have features unobservable to the researcher that have both determined their network structure and position as well as their outcomes. This suggests that an examination of a periphery actor's connection to a core/periphery bridge and their outcomes may conflate the role the network plays with innate qualities of the individual (Manski 1993; Jackson and Wolinsky 1996; Goldsmith-Pinkham and Imbens 2013). Furthermore, network studies generally examine connections that already exist. Therefore identifying a comparable control group is extremely difficult, as those with connections are likely different from those without connections.

To overcome these empirical challenges, I exploit variation in the timing of the formation of a core/periphery bridge through evaluating the impact of the return home of a foreign trained scientist to developing countries on outcomes of non-migrants affiliated with the institutions they return to. Scientists returning from developed countries back home to developing countries can be considered insiders whose return home results in the formation a core/periphery bridge for peripheral non-migrant scientists.

Specifically, I study the effect of the return home of 112 HIV researchers trained in top universities in the United States under the National Institute of Health Fogarty AIDS International Training and Research Program between 1988-2014. I construct a panel dataset of 1,657 non-migrant African scientists who are affected by these return events in that they are working in related fields in the institution to which the American-trained scientist returned. Matching with scientists from other institutions in Africa that do not receive a returning trainee, I am able to control for career, field and temporal trends in research output. Difference-in-differences regressions compare within scientist changes in publication outcomes of active researchers in institutions following the return of an American-trained researcher with changes in publication outcomes of observably similar researchers in other African institutions.

The results reveal increases in the rate at which non-migrant scientists collaborate with scientists from the American training institution of the returning scientist following the return event. Non-migrants also increase citation rates to scientists based in the American training institution of the returning scientist following the return event. Furthermore, they experience a persistent increase in publication output following the arrival home of an American-trained scientist, particularly in HIV research. The effect is most pronounced for non-migrants who are not connected to developed country scientists prior to the return event. The findings support the idea that a returning scientist forming a core/periphery bridge benefits periphery actors. Potential mechanisms are explored, and evidence is found in support of two possible drivers of the effect. Non-migrants can both access knowledge that the return migrant has access to, as well as the connections of the return migrant.

The rest of the paper proceeds as follows. Section 2 discusses the theoretical framework. Section 3 describes empirical setting, the National Institute of Health Fogarty AIDS training and research program. Section 4 describes the data and provides some descriptive statistics. Section 5 presents the results and section 6 concludes and outlines implications of the findings.

3.2 Theoretical Framework

Can Returning Scientists Broker?

Skilled migrants moving back to developing countries would seem to be in an ideal position to bridge developed and developing country networks and transfer connections, knowledge and resources back home. In her book 'The New Argonauts', Saxenian (2006) places considerable weight on this phenomenon amongst entrepreneurs moving back to Taiwan following experience in Silicon Valley:

> *'But these highly skilled emigrants are now increasingly transforming the brain drain into "brain circulation" by returning home to establish business relationships or start new companies while maintaining their social and professional ties to the U.S.'*

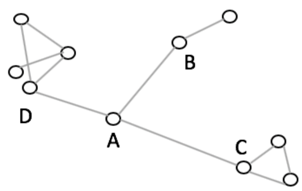
Yet despite this potential for return migrants to bring connections, knowledge and resources back home, recent work suggests the difficulties in this brokerage role amongst scientists. Kahn and MacGarvie (2016a) study the movement home of Fulbright scholars following American training and find that the returnees to developing countries experience a significant decline in their productivity as compared to carefully matched scientists who remain in the United States. They ascribe this to distance from resources and knowledge production which can limit the ability of scientists to fully transfer connections, knowledge and resources back to developing countries. Accordingly, Gibson and McKenzie (2014) find a productivity decline for scientists trained abroad returning to the Pacific Islands, even after controlling for negative selection of the returnees. Together these results suggest that the benefits from being a broker between foreign and home locations of returning scientists are limited and that the net effect is negative.

Second-Hand Brokerage

But these discouraging findings may reflect a conception of brokerage that is overly narrow. In particular, there is reason to think that returnee scientists are especially likely to facilitate what Burt (2007) called 'second-hand-brokerage'.

Burt introduced the concept of second-hand brokerage to capture the possibility that actors associated with a broker might be able to effectively access a broker's connections, knowledge and resources. Figure 1 illustrates this concept, where nodes B, C and D are 'at risk' of accessing A's connections, knowledge and resources - or to become 'second-hand brokers'.

Figure 1: Illustration of Potential for Second-Hand Brokerage



There are two possible mechanisms whereby second-hand brokers can benefit: (1) accessing knowledge through the broker, (2) accessing the connections of the broker.

Accessing Knowledge The first of the two mechanisms assumes that knowledge can flow freely through indirect ties (Granovetter 1973; Watts and Strogatz 1998), and therefore second-hand brokers have access to the knowledge the broker has access to. As an example, node D in Figure 1 could access information from B and C, via A.

Accessing Connections The second mechanism assumes that networks are not static. In particular, brokerage positions can close over time (Burt 2002) as brokers connect previously unconnected actors (Obstfeld 2005). The broker can close the gap between disconnected actors, providing endorsements, knowledge about each other or introductions. This closing of the gap can result in actors on each side of the broker to form direct relationships with each other (or closing the triad). As an example, node D in Figure 1 could create relationships with the connections of broker A in a second time period. This mechanism subsumes the first mechanism as both result in access to the knowledge that broker A originally had access to.

Together, these mechanisms suggest that association with a broker can enhance access to connections, knowledge and resources that may lead to subsequent improvements in performance. In practice though, evidence on second-hand brokerage is limited (Burt 2007). Burt has offered two reasons why sharing connections, knowledge and resources might be difficult. First, if the potential second-hand broker can build their own brokerage position, then there is no additional advantage to second-hand brokerage. If a potential second-hand broker can observe the structure of the network and wants to build relationships such that they are a broker, second-hand brokerage is redundant. Second, the extent to which knowledge is ‘sticky’, or hard to communicate, can overwhelm the possibility of second-hand brokerage. Another possible reason why second-hand brokerage might be limited is that the broker themselves requires an incentive to share their knowledge and connections with potential second-hand brokers. To the extent that the broker can extract rents from their position bridging disparate parts of a system, it is naïve to think that they would give that up and share their position with potential second-hand brokers without receiving something in return (Reagans and Zuckerman 2008).

Whilst each of the foregoing three considerations imply that second-hand-brokerage will have limited benefits, they also imply three contextual factors that can enhance its value. First, whereas it might often be true that potential second-hand brokers can build their own brokerage position, this is sometimes not the case. Outsiders in a system face barriers in taking advantage of their own network position due to a lack of legitimacy, and subsequent mistrust, amongst the community.[8] However, being associated with an insider, or a sponsor can allow them to take advantage of borrowed social capital (Burt 2010). Second, whereas key knowledge and knowhow is often local, there exist settings where organizations and routines exist to transfer knowledge. For example, within organizations (Burt 2010; Choudhury 2015) or science (Mohnen 2016). And third, when the broker has an incentive to share knowledge and connections or can take credit for the success of someone else, or when they are constrained to take full advantage of their brokering position, sharing of knowledge and connections might take place. Examples of such settings

[8] The idea that insiders have some privileged access to knowledge within the community of knowledge producers was formalised in Merton (1972), who attributes differences in status to trust.

are those between mentors and their protégé (Burt 2007; Choudhury 2015), or within teams (Mohnen 2016).

Burt (2010) provides a case of female managers in an electronics manufacturing firm to illustrate the potential benefit to second-hand brokerage in a setting that meets the three conditions described above. Women, who he considers outsiders here, in the firm who had male sponsors who were invested in their careers and central to other individuals' networks were promoted sooner. Burt attributes this difference to male sponsors sharing their social capital. This sponsorship, or endorsement, from an insider reduced legitimacy problems that the women in the firm originally had and enabled them to take advantage of the connections made as a result.

Further evidence supports this interpretation of second-hand brokerage for outsiders. Entrepreneurial firms affiliated with more prominent exchange partners have a faster rate of initial public offering and earn greater valuations at IPO than similar firms without such connections (Stuart et al 1999). Graduate students with prominent advisors tend to find better initial job placements than comparable students without such prominent advisors (Long et al 1979), and West African scientists experiencing a random shock to their ties with developed country scientists as a result of the ebola outbreak experience improvements in their publication output (Fry 2019).

Although not framed in terms of second-hand brokerage, two recent studies provide support that association with a brokering individual in knowledge production can impact outcomes. Choudhury (2015) finds that returnee managers moving back to India within a multinational enterprise exert positive benefits onto their R&D employees through forming a bridge between knowledge generated in the headquarters in the United States to R&D employees located in India. Mohnen (2016) finds that the impact of the death of star scientists on their collaborators is more negative if the star is a broker in the network, and if the collaborator is younger or more isolated.

While this prior literature has established an empirical relationship between second-hand brokerage and performance of outsiders, the scope of this effect, particularly in global science, may still be quite limited. Within an organizational or team context there are routines and incentives to transfer knowledge and motivation to see others succeed.

Amongst scientists there aren't always natural routines for sharing knowledge, and there is a great deal of competition between individuals. That being said, science is highly codified and there are incentives for scientists to distribute their knowledge, and so one might expect knowledge to flow through the network more easily than in other settings. Furthermore, there are some relationships amongst scientists - for example, the mentor-protégé relationship - in which incentives do exist to share knowledge and connections with the outsider. Together with the existence of insiders and outsiders, global science provides an interesting case in which to explore the possibilities of second-hand brokerage.

Core/Periphery Bridges in Global Science

Global science demonstrates a network structure with insiders and outsiders crudely classified by their geographic location. Classic accounts of global science networks describe the structure as core/periphery, with the majority of citations, collaborations, publications and patents occurring in more developed countries in the world, with the gap widening over time (Wagner and Leydesdorff 2005; Leydesdorff and Wagner 2008; Hwang 2008; Zelnio 2012). Peripheral actors in this system suffer from a lack of access to resources, central knowledge and other benefits.

In light of the discussion above on the relevance of second-hand brokerage for insiders and outsiders in a system, a core scientist forming a bridge to periphery scientists (a "core/periphery bridge") provides an opportunity to test ideas about sponsorship. Periphery scientists associated with the bridge can access a core scientist's knowledge, connections and resources. Additional features of the scientific setting give further insight as why the presence of a core/periphery bridge could result in improved access, and as to the dominant mechanism(s) driving any observed changes.

Given that scientific knowledge is codified and incentives exist to distribute it widely (Stephan 1996; DasGupta and David 1987), the presence of a core/periphery bridge should facilitate flows of knowledge across disparate parts of the system. Additionally, scientists act as mentors to others as a way to build a legacy. With the assumption that scientists would like to leave behind a legacy (or that they are helpful), periphery scientists associated with the core/periphery bridge could access their knowledge and connections. In light of this, and given previous evidence demonstrating the positive impact of access to the core on periphery scientist performance (Fry 2019), performance of the periphery scientist associated with a core/periphery bridge should be positively impacted.

To this point, the discussion has assumed that core/periphery bridges are built somehow - an assumption to which I will return momentarily. For now, assuming that such a bridge is built, this implication follows: *periphery scientists associated with a core/periphery bridge can access connections, knowledge and resources of the broker, resulting in improved performance of these scientists.*

Returning Scientists as Core/Periphery Bridges

But can core/periphery bridges be built? And if they are, will they have a causal impact? These implications are challenging to test because the network surrounding an individual is rarely randomly determined (Manski 1993; Jackson and Wolinsky 1996; Goldsmith-Pinkham and Imbens 2013). Actors in a network have qualities that determine both their network and their outcomes, conflating the role of the network in outcomes with innate qualities. Furthermore, because we generally observe actors with a given network, it is very difficult to define a comparable control group, as by the time they are observed they have already experienced different paths.

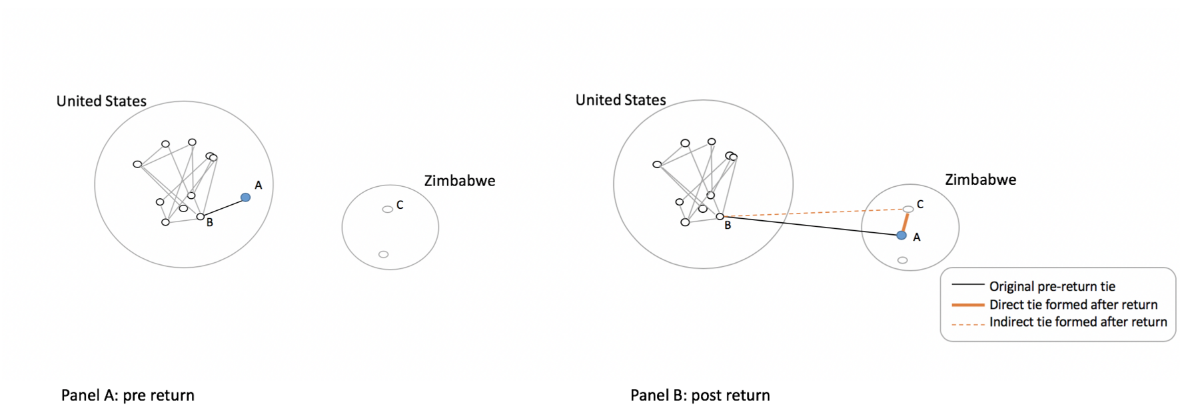
I exploit a setting in which I can isolate the timing of the formation of a core/periphery bridge, and then examine the outcomes accruing to the peripheral scientist associated with the bridge before and after the shock. This longitudinal contrast removes omitted variable bias that cross-sectional comparisons face. However, if one expects that the timing of the formation of a core/periphery bridge is endogenous to expected performance improvements, then this difference-in-differences estimate could be biased. That is, connections to particular periphery scientists could be made if core scientists expect that those periphery scientists will perform well. If the best selected scientists are also those who are subject to the formation of a core/periphery bridge, estimation of the change could reflect positive selection as opposed to a causal effect of the core/periphery bridge. To remedy this problem, I pair each treated scientist with a control scientist who exhibits almost-identical performance prior to the potential formation of a core/periphery bridge, and analyse the data at the individual level of analysis in a difference-in-difference framework. This provides a reliable way to evaluate the effect of the formation of a core/periphery bridge.

The return home of scientists to developing countries provides an ideal setting for this study. Past research has demonstrated that mobile scientists take their networks with them (Azoulay 2012; Scellato et al 2015). In addition, co-location increases the probability of forming a social relationship (Boudreau et al 2017; Catalini 2018), particularly within the same institution (Azoulay et al 2012). Together, this suggests that a scientist who returns to a developing country from a developed one occupies a position in the core/periphery network whereby they are considered an insider, and form a bridge between two disparate parts of the system. For periphery non-migrant scientists, the return home of a scientist from more developed countries to their institution implies the formation of a core/periphery bridge.

To illustrate this idea, consider Figure 2. This figure, in which nodes represent scientists and ties their connections presents a core/periphery network structure. The darker nodes at the center of the figure (individual B and A) represent core scientists, and the lighter colored node (individual C) represent periphery scientists (without loss of generality, core scientists are those based in the United States and periphery scientists are those based in Zimbabwe). Panel A represents the pre-return state, where periphery scientists are unconnected with the core. Panel B represents the network after the return home of the scientist (scientist a) from the United States. Taking her connection to core scientist B with her, and forming a new relationship with periphery scientist C, the return home of scientist A to Zimbabwe from the United States provides a bridge between the core and periphery. Specifically - scientist C, the second-hand broker, who is connected to returnee A can share the returnee's access to knowledge, connections and resources in the core.

The return home of a scientist from a more developed country implies a shift in opportunity to periphery non-migrant scientists in their institution to access central connections, knowledge and resources. As such, it provides a lens to examine how performance is affected by second-hand brokerage. Specifically, the presence of a core/periphery bridge in the form of a returning scientist may result in improved access to knowledge, collabora-

Figure 2: Core/Periphery Bridge Formation



tors and resources from the core, which in turn may result in improvements in publication outcomes, particularly in the research area of the returning scientist.

Whilst reflecting on how the return home of American-trained scientists to developing countries have impacted their home countries, one American-based mentor I interviewed stated:

> *'...there is an impact of the returning trainee. One is the contribution of that individual is from teaching and publishing, and one is the bridging....this bridging is particularly important when you have minimal resources for research.'*

Another American-based mentor I interviewed described how returnees bridge his own institution with their network back at home:

> *'Just last year we were give some money by a private corporation to do some capacity building in low income countries – I was looking at who could we recruit to come here for training, and I relied on my network who had been here [in the United States] who are back in country...'*

As described earlier, there are two main ways that the return home of a scientist from developed to developing countries can impact non-migrant scientists in developing countries. The first possibility is that the non-migrant can access the knowledge that the return migrant has access to, subsequently affecting the performance of the non-migrant (Hoenen and Kolympiris 2018; Ganguli 2015; Kahn and MacGarvie 2016b; Singh 2005; Agrawal et al 2006). The second is that in addition to access to their knowledge, the non-migrant can access the connections of the returning scientist, which can affect subsequent performance due to new collaborative relationships (Azoulay et al 2010; Wuchty et al 2007) and/or benefits from elevated status (Azoulay et al 2013).

Although these mechanisms are very hard to separate, because by definition the existence of the second mechanism obscures the existence of the first within an individual, direct measurement of knowledge flows and collaborations (which imply the existence of each mechanism, respectively) and contextual factors may help to distinguish them. There is reason to think that access to connections and subsequent new relationships, which is based on reputation inferred from association with the broker, is particularly important for those with which there is the greatest uncertainty (Stuart et al 1999). Furthermore, ceiling effects to reputation have been found in sciences (Azoulay et al 2013). Both of these imply that if the non-migrants accessing the connections of the return migrant, non-migrant scientists with previous connections to central actors - who already provide a signal of quality to the community - would not benefit as much from the arrival home of a return migrant from more developed countries. As such, it will be important to examine which non-migrant scientists are most impacted by the return migrant.

The remainder of the paper tests these propositions through examining a program that systematically supports low income country scientists to study in the United States and return home following their studies. The next section provides details of the program.

3.3 Empirical Setting

The empirical setting for this paper is that of life sciences research in African institutions, and the return home of African scientists who took part in long-term training in the United States supported by the National Institute of Health (NIH) Fogarty International Center AIDS International Training and Research Program (FIC AITRP).

AIDS International Training and Research Program

Established in 1968, the NIH Fogarty International Center (FIC) funds around 500 research and training projects across 100 American universities, and 120 countries. With a budget of just over USD $75 million in 2018, they boast having contributed towards major advances in global health and Low and Medium Income Country (LMIC) scientific capacity development.

The flagship program when it comes to human scientific capacity building LMICs is the AIDS International Training and Research Program (AITRP) (now known as the HIV Research Training Program). Started in 1988, this program was developed in response to the HIV epidemic and the perceived need for strengthened scientific capacity in AIDS endemic countries around the world.

> *'But it really changed with the AIDS epidemic, and the realization that to address this particular epidemic we had to change our style of conducting research internationally. We had to overtly move away from the 'colonial' research, or the 'parachute' approach, and really get into collaborative research and capacity building on site.'* Gerald Keusch, MD, Director of FIC 1998-2003

AITRP provides grants to principal investigators (PIs) in American universities to work with LMIC sites (universities, hospitals and research centers) in strengthening their human capacity through training of scientists, clinicians and allied health workers in research. This training is offered as short or long-term (graduate and non-degree studies, generally over 6 months in duration) programs, usually with a combination of American and field site location. The American-based PIs apply for, and receive, grants in five year cycles, of around USD $500,000 a year, renewable upon re-competition. While the first cycle in 1988 involved eight American institutions, this has now expanded to include around thirty American institutions offering a variety of HIV related research (with TB added in later on) training programs across the world. The American universities involved in the program are some of the leading educational and research institutes: including Harvard University, Johns Hopkins University, Brown University and many more. Between 1988 and 2010 FIC claims to have contributed towards training 1,559 LMIC researchers in long-term AITRP programs with a cost of just over USD $200 million under the AITRP.

The United States based PIs are at liberty to design the training program, across any HIV/AIDS related fields. Most run a variety of short-term programs (workshops, summer courses usually at the LMIC site), although the long-term, degree level (Ph.D, masters) as well as non-degree (including post-doctoral), training for individuals from the LMIC site is generally the focus of the program. In the earlier days of the program, many of the longer term trainees came from institutions in LMICs other than the main site described in the grant.

FIC specifies in the grant announcements that the long-term trainees should be given incentives to go back to their country of origin. A survey carried out by FIC in 2002 found that a return rate of over 80% at that time.[9] This return home is not necessarily to take up a research position, or to the LMIC site involved in the program, moreover FIC prides itself on graduating trainees assuming high level positions in government and multilateral organizations. Incentives to return home include 'sandwich training', strategic selection of trainees, re-entry funding, visa restrictions and formal return agreements and contracts with their training institution.

The returning trainees studied in this paper are those African scientists who participated in long-term FIC AITRP supported training at American institutions between 1988 and 2014 inclusive. The FIC AITRP program was one of the first programs around the world to engage researchers from Africa in systematic training in frontier research, and has boasted as contributing towards enormous achievements in terms of research and public health outcomes, HIV and otherwise, in African countries. As just one example, FIC AITRP trainees and collaborators were responsible for the landmark 2011 study HPTN 052[10] which revealed the personal and public health benefits of early treatment and led to treatment guidelines on treatment as prevention. At the time, Executive Director of UNAIDS, Michel Sidibé, described the results of HPTN 052 as a *'breakthrough'* and *'a serious game changer [that]... will drive the prevention revolution forward.'*[11]

[9] https://www.fic.nih.gov/News/GlobalHealthMatters/july-august-2012/Pages/hiv-aids-aitrp-program-anniversary.aspx last accessed 10-8-19

[10] https://www.annalsofglobalhealth.org/articles/10.5334/aogh.2432/

[11] https://www.unaids.org/en/resources/presscentre/

3.4 Data and Sample Characteristics

This section provides a detailed description of the process through which the data used in the econometric analysis are assembled. I describe (1) the sample of returning African scientists trained in the United States; (2) the sample of non-migrant African scientists affected by these returns, and the set corresponding control scientists to which they will be compared; (3) outcome variables used in the study. I also present relevant descriptive statistics.

Sample of Returning African Scientists Trained in the United States

Names of African trainees participating in the FIC AITRP program are gathered directly from the records of American institutions involved. The 20 (out of 29 total universities and 2 training hospitals identified receiving AITRP grants) American universities involved in the AITRP with African trainees are contacted, and 14 of these universities provide data on their long-term African trainees. The universities provide the names, home country, degree (if any) and year(s) of training for their long-term African trainees between 1988 and 2014.[12] The names of the trainees are then matched with publication data, if any, using the Elsevier Scopus database, and information on institution of return for each trainee is gathered based on publication affiliations post graduation. Resumes of trainees are gathered using a combination of internet searches and direct email correspondence to ascertain the returnee's role in the institution of return, as well as institutions of non-publishing returnees.

The sample of trainees contains 421 African researchers who took part in long-term training in the United States over the full time period (1988-2014). Trainees are from a range of African countries, with each American university hosting trainees from a variety of countries. I identify those trainees who returned home to Africa following graduation using affiliation information from publication data and information on graduation year. The trainee is considered to have returned home if they published following the year of graduation from the training program and their affiliation was in their home country. Out of 284 returnees with a publication record with affiliation details following graduation,

[12]Some of the universities state that their records of trainees in the 1990s is poor - and so it is possible that the sample is biased towards more recent time periods.

242 trainees (85% of all those continuing to publish, or 57% of all trainees) in the sample returned home, while 34 of them remained in a developed country (mostly the United States), and 8 moved to African countries other than their home country. Because some returnees are affiliated with more than one institution in their home country after training, I identify 316 unique return events across institutions in 15 African countries between 1988 and 2014 (Figure 3; Table 1). As the timing of the return home cannot be ascertained precisely from publication records I use the year of graduation from the American program as a proxy year for return events.

Sample of Treated, Non-Migrant African Scientists

I measure the impact of exposure to American-trained scientists returning to African institutions on local scientists. Therefore I focus on scientists working in the institutions in Africa at the time of return of the FIC AITRP trainee. Those scientists affiliated with the institution that the FIC AITRP trainee moves (back) to (publishing within the 3 years before the FIC AITRP trainee graduates, with over 75% of publications in those 3 years affiliated with that African institution), and working in HIV related research (i.e. published at least one HIV related publication in the 3 years prior) are considered treated by the return event of the American-trained scientist.

In order to identify those scientists treated by the return event of American-trained scientists I use publication data in the Elsevier Scopus database to generate a sample of scientists and associated publication history affiliated with each African institution between 1988 and 2014.

The use of publication data in studies of this type (namely, in generating a plausible set of scientists in a particular location associated with their full publication record) comes with two major challenges. First, generating a full scientist level publication record is complicated by the fact that scientists may have common names (for example, Smith J), therefore it is difficult to determine which Smith J published which paper, or a single scientist may go by more than one version of a name. Second, understanding where scientists are located given that an affiliation in a publication may not accurately represent the full-time location of a scientist, and that the scientist needs to publish in order for the researcher to see their affiliation — which for the African scientists is not always the case in each year. Fortunately, the first issue is resolved using the Elsevier Scopus publication database which provides an author identifier for each author in every publication contained in the database. The author identifier is developed using an algorithm that incorporates scientist name, coauthors and topic type and allows for scientists to change affiliations across publications. The second issue is resolved using a rule of thumb — if a scientist classifies her affiliation as being in a certain country in over 75% of her publications over a defined, multi year period, she is considered as being based in that institution in that time period. The first year of treatment is considered as the treatment in the main analysis. A single scientist can be treated by multiple returnees coming from multiple American institutions in the same year.

Carrying out this procedure gives 1,740 scientists treated at some point during their career by a return event of a FIC AITRP trainee. As the first year of treatment is the main event considered, several of the returnees drop out of the sample (those who were the second or third trainee to return to the same institution within the career of a treated non-migrant), leaving 112 returning FIC AITRP trainees considered in the main analysis. This gives a median of 11 scientists (mean 19) treated by the return of a single FIC AITRP trainee (Figure 4). I match the list of treated non-migrant African scientists with their full publication record and generate variables of publication rate, collaboration outcomes and content of the research. The complete list of references used in each scientist's publications are also gathered from Elsevier Scopus database.[13]

Sample of Control Scientists

In order to identify the effect of return of an American-trained scientist I could examine changes in non-migrant African scientist's output after the return event, relative to before the return. However, because the return event is correlated mechanically with career age, and calendar year, the specifications must include life cycle and period effects (Levin and Stephan 1991). The control group that pins down the counterfactual age and calendar time effects are those scientists who never experience the return of a sample FIC AITRP trainee in their institution.

Publication data in the Elsevier Scopus database is again use to generate a sample

[13] using code developed by Rose and Kitchin (2019)

of scientists affiliated with institutions in countries that were at some point involved in the FIC AITRP (to ensure that scientists included are from countries that are similarly connected with the United States and equally economically and politically stable). The control scientists are culled from this universe of African scientists who are affiliated with institutions that did not receive a returning FIC AITRP trainee in the time period of the scientist's career. The control scientists are chosen using a coarsened exact matching procedure so that their involvement with American-based scientists (from American FIC AITRP training institutions and any other American institutions), type of research (HIV or otherwise), and publication rate matches that of the treated scientists at the counterfactual date of treatment (Appendix A provides more details on construction of the control group). Combining the treated and control samples allows me to estimate the effect of the return of an American-trained scientist in a difference-in-differences framework.

In addition to quantitative data on publication outcomes, I conducted 16 interviews with NIH FIC staff, American-based scientists and administrative staff involved in the FIC AITRP grants, as well as with African trainees and other scientists in institutions receiving returnees. These interviews were carried out on the telephone, or in person where possible. They ranged from 30 minutes to 2 hours, and the primary purpose of the interviews is to illuminate mechanisms of impact of the FIC AITRP and returning trainees.

Measurement

I use a variety of different measures all of which are generated using publication records of the non-migrant African scientists. First, I generate a number of publication count measures to identify changes in publication rate of scientists. Second, I measure collaboration patterns. Finally, I measure knowledge flows between returning trainees and their American-based networks using a variety of approaches.

Measurements can be divided into more general measurements (such as publication counts and collaborations with any American-based scientist), and measurements that incorporate the returnee or the returnee's American training institution (such as rate of collaboration with the returning scientist or returnee's American training institution). As the treated scientists are matched one to one with control scientists in the CEM matching procedure based on pre-treatment (or counterfactual year of treatment) variables, a counterfactual year of return, returnee and returnee's American training institution is given to each control scientist. The control scientists are then assigned to their matched treated scientist's returnee and associated American training institution, and measurements based on the specific (counterfactual) returnee can be generated for both treated and control scientists. For those treated and control scientists who experience more than one (counterfactual) returnee in a given year, outcomes are measured for each returnee, and associated American training institution, and the maximum value is taken.

Rate of Publication

Measures corresponding to the rate of publication include the number of publications each year a scientist is an author on, and an additional measure weighting each publication by its journal impact factor — a measure of the frequency with which the average article in a journal has been cited in a particular year. Key word searches of the title, abstract and keywords in each publication in a given year for a sample scientist gives publication outcomes in a given research area. The probability that at least one of these publications contains keywords associated with HIV related research is measured.

Collaboration Rates

Collaboration patterns of non-migrant African scientists is measured using co-authoring patterns in publication data. Using author written parsing script I extract coauthor names and affiliations from the scientist's publication record to generate collaboration counts (both absolute and dummy indicator) across a variety of groups.

Measures of collaboration rates with any American-based scientist, or any scientist affiliated with an American training institution involved in the FIC AITRP are generated. In addition to collaboration rates with these two groups of scientists, more specific measures based on collaboration rates with the returning trainee and scientists from their American training institution are generated.

Knowledge Flows

I measure knowledge flows between the sample scientist and the returning trainee, the returnee's American training institution and any American-based scientist. Following a long line of research, I use publication to publication citations to measure knowledge flows among scientists (Jaffe et al 1993; Jaffe and Trajtenberg 1999). Taking the non-migrant scientist's full list of references used in their publications, I measure the extent to which they cite publications authored by the returning trainee, the returning trainee's American institution, as well as any American-based scientist[14] in each year before and after the return event.

Measurements for control scientists are made in a symmetric way, using the counterfactual return event date, returning trainee and American training institution.

Descriptive Statistics

I identify 1,740 distinct scientists who are affiliated with an African institution at the time of the return of a FIC AITRP trainee. The matching procedure identifies a control scientist for 1,657 (95%) of the treated scientists, treated by 112 unique FIC AITRP trainees.

FIC AITRP Trainees

Figure 5 illustrates the publication trends of the 288 FIC AITRP trainees who have a publication record following graduation. The differential trends of trainees known to return home versus those remaining in developed countries shown in the figure is most likely due to a selection effect resulting in different types of trainees not returning home. One important point to note from the figure (panel d) is that the returning trainees continue their collaborative relationship with their American training institution following graduation.

Descriptive statistics for the sample of 112 FIC AITRP trainees used in the main analysis is provided in Table 2. The average returnee graduated from their fellowship in 2004. 65% of returnees returned to institutions with broader institutional programs,[15]

[14] Just publications authored by returnee and American scientists in the 5 years prior to the return event are considered in order to account for the possibility that the publication themselves may be influenced by the return (as per Azoulay et al 2012).

[15] FIC AITRP grants pre-specify at least one institution in Africa in which the United States based grant holder carries out a variety of shorter term programs in.

and almost 80% returned to institutions in which there had been previous returnees from the FIC AITRP program. In the five years following return home, most of the returning scientists publish some research (with an average of 5.5 publications in the five years after return), particularly in HIV. 63% of returnees continue to co-author with scientists from their American training institution, and 75% co-author publications with scientists from the institution they return to.

Non-Migrant African Scientists

The descriptive statistics in Table 3 pertain to the set of 2 $\times$ 1,657 = 3,314 matched treated and control scientists. The covariates of interest are measured prior to the return of the trainee (or counterfactual). A few of the covariates are balanced between treated and control scientists by virtue of the CEM procedure — for instance, the career age at the time of return. However, the observed balance in other statistics, such as the five year stock of number of American coauthored publications, number of any FIC AITRP American institution co-authored publications, and number of HIV publications at time of (counterfactual) return is not guaranteed.

While publication outcomes are well matched at baseline, there are differences between the mean likelihood of treated and control scientists to co-author with the returnee, and the returnee's training institution. This is consistent with the view that moves are not random, and that there are some people in destination locations who a mover is already connected to. Because these baseline differences make it difficult to ascertain co-authoring behavior in the absence of the return event, those non-migrant scientists who have co-authored with the trainee or the trainee's American training institution are removed from the analysis in robustness checks.

The estimation sample includes observations 5 years before and after the return event (or counterfactual). The result is a balanced panel dataset with 36,454 scientist-year observations.

3.5 Results

Econometric Framework

In order to identify the effect of the return of an American-trained scientist, I compare a non-migrant African scientist's outcomes after the return of the American-trained scientist to their institution relative to before, using a scientist fixed effect specification. The estimating equation (equation 1) relates non-migrant African scientist i's outcomes in year t to the return of a FIC AITRP trainee in their affiliated institution.

$$\begin{aligned} E\left[y_{it}|X_{it}\right] = exp\Big[&\beta_0 + \beta_1 AFTER\_RETURN_{it} \\ &+ \beta_2 AFTER\_RETURN_{it} \times RETURN\_INSTITUTION_i \\ &+ f(AGE_{it}) + \delta_t + \gamma_i\Big] \end{aligned} \tag{1}$$

Where y is the outcome measure, AFTER_RETURN denotes an indicator variable that switches to one the year after the FIC AITRP trainee returns to the home country, RETURN_INSTITUTION is an indicator variable that switches to one if the sample scientist is treated. f(AGE) corresponds to a flexible function of the non-migrant African scientist's career age, with calendar year fixed effects and non-migrant African scientist fixed effects. Standard errors are clustered at the level of the institution.

The majority of the dependent variables of interest are skewed and non-negative (Figure 6 illustrates the distribution of publications the year of the return (or counterfactual)). Due to the large number of zero's in the dataset, most of the specifications are estimated in two ways; first using an inverse hyperbolic sine ordinary least square estimate. And second, outcomes are converted into dummy outcomes (given a 1 if they happen at all in a given year), and a second set of estimates use a linear probability model.

The Impact of a Returnee on Non-Migrant African Scientist Performance

Table 4 presents the core results estimating the specification presented in equation 1. These provide strong support for the expectations of the paper, that return home of an American-trained scientist results in an increase in performance of non-migrant scientists as well as a relative increase in performance in the field of the returning scientist. Overall, I find that non-migrant scientists increase their rate of publications following the return home of a FIC AITRP trainee, as indicated by the estimates for AFTER_RETURN × RETURN_INSTITUTION being positive and statistically significant (column 1, 2, 3, 7). I find a sizable and significant 6% increase in the annual number of publications of a non-migrant scientist following the return of an American-trained scientist, as compared with a scientist not subject to the return of an American-trained scientist (column 1). To verify that this isn't driven by an increase of publications in low quality journals, column 3 measures the change in publications weighted by their journal impact factor. The significant increase for the treated group as compared to the control group suggests that scientists increase both quantity and quality of publications following the return event. The increases in rate of publication are due mostly to increases in HIV related research (column 4, 6). Column 6 shows that treated scientists are 3.6 percentage points more likely to publish in HIV related research following the return event. With the average probability of publishing in HIV related research of 0.32, this post return increase is around a 10% increase on the mean. However, the non-migrants do not experience a significant increase in the probability that they have first, or last, authored publications (column 7 and 8), which raises questions on the role of the African scientists on projects, and the possibility that they are a 'long-arm' of the American labs, a concern of which I return to in the discussion.

I explore the dynamics of these effects in Figure 7. I estimate a specification in which the treatment effect is interacted with a set of indicator variables corresponding to a particular year relative to the trainee return. Three points worth noting from these figures: (1) effects do not appear to be transitory; (2) although the results are noisy due to the small sample, there does not appear to be a pre-trend; and (3) the impact appears immediate following the return event. At first glance this seems confusing due to an assumed lag between the start of projects and publication time. There are two potential reasons for this observations. Scientists could be joining projects near to the end of the project (which if they are primarily co-ordinating field work it is entirely possible they are not engaged in the grant writing and design stage of the project). The other possibility is that the returnee is connecting non-migrants and American-based scientists prior to their return. The program studied encourages trainees to return to their home countries during the program to carry out field work, therefore it is possible that in the last stages of the program they are already acting as a broker. The rest of the paper explores the mechanisms by which a return migrant impacts non-migrant performance.

Return Migrant as a Core/Periphery Bridge

Figure 8 illustrates initial support that the return migrant provides a core/periphery bridge between institutions in Africa and the United States. Depicting the institutional collaborative network both before the FIC AITRP program begins (panel a), and after (panel b), the figure shows that following the program, treated institutions (lighter gray circles) become more central to the full network of African and American institutions, as compared to the control institutions (white circles). The figure also illustrates that all of the African institutions in the sample become more connected over the full time period. This fact further necessitates the use of control scientists in the sample to account for this trend. Subsequent evidence that the return migrant is providing a core/periphery bridge is explored through unpacking each of the proposed mechanisms by which second-hand brokerage can operate.

Accessing Knowledge

I first test whether knowledge flows from the returnee's network in the United States increase after the returnee's arrival. This would be suggestive of a bridge being formed, and the broker sharing access to their knowledge with their affiliates.

I find that treated scientists tend to cite publications authored by scientists from the returnee's training institution (or counterfactual) more following the return event (Table 5 column 2, 3, 4). With an increased probability of citing the American training institution of the returning scientist of 2.2 percentage points, this gives an economically significant 29% increase over the mean. Interestingly, this increase in citations to the returnee's training institution is even observed for publications that are not coauthored with the training institution (column 5). This implies that the non-migrants are learning about the research taking place in the training institution either in coauthored publications and carrying it over, or directly.

Accessing Connections

Second, I assess whether association with a broker allows actors to access their connections through measuring non-migrant African scientist's collaboration rates with American-based scientists that the returnee is connected to.

Figure 9 depicts the collaboration rates of treated scientists with various different groups of American scientists. Panel B illustrates that the treated scientists are more likely to coauthor a publication with an American-based scientist affiliated with a FIC AITRP training institution after the return event. Table 6 provides the regression counterpart to Figure 9, and columns 1, 2 and 3 illustrate that non-migrant scientists are more likely to collaborate with scientists from the United States, in particular those from the training institution of the returnee (or counterfactual). Treated scientists are 33% more likely to publish with the American training institution following the return event (column 4). These are mostly new collaborators for these non-migrants (Table 6 column 5, 6). This provides supporting evidence that under certain conditions, actors associated with a broker can access their connections.

As discussed in Section 2, one would expect non-migrants with fewer connections to the core prior to the return event to experience greater improvements to their performance if the return migrant allows the non-migrant to share their connections through some kind of sponsorship. Tables 7 explores heterogeneity in the effect of the returnee through separating the treated and control scientists into three groups based on their network prior to the return: 1. those who publish with the returnee's American training institution prior to the return, 2. those who publish with OECD based scientists in over 75% of their publications prior to the return,[16] and 3. the remainder. The same difference-in-difference regression is run on the three groups separately. As seen in columns (3) and (6), the greatest impact of the returnee is felt by those scientists less well connected with OECD scientists prior to the return. This also serves as a robustness check as the movement of the returning scientist to an institute is likely to be endogenous to the scientists in the institute who were already within the same close network. The fact that the greatest effect is not experienced by those non-migrants with a collaborative history with the returnee's American training institution (columns 1 and 4) is comforting.

Another piece of supporting evidence that the impact of a return event is at least in part due to sharing the returnee's connections is the differential impact of a returnee according to their role back home. One the one hand, if the non-migrants benefit predominantly from accessing the knowledge of the returnee, it might be expected that the benefits are greater when the returnee is an active scientist. On the other hand, if the non-migrants benefit predominantly from accessing connections of the returnee, it might be expected that the benefits are greater when the returnee has more of an administrative, or outward facing role in their institution. Interviews suggest that the latter is true, and the following quotation from an interview with an American-based PI involved in the FIC AITRP confirms how the role of a returnee can influence their impact:

> *'One of my trainees was chairman of the School of Medicine.... He had a credible skill set, but he was not able to put his skill set to use because he was*

[16] I choose this definition that requires a scientist to have published 75% of their publications pre-return event (or counterfactual) with OECD based scientists in order to distinguish between those who are truly embedded in OECD scientific networks, and those who have incidentally participated in collaborative research projects. Almost 80% of sample scientists have coauthored with OECD scientists at some point prior to return events (or counterfactual) and the median proportion of OECD coauthored publications across scientists in the five years prior to return events is 0.56. However, field work highlighted variation in the approach of African scientists to collaborative field work. Some scientists consider themselves as working in the same lab (albeit remotely) as OECD based scientists, sharing equipment and funding with frequent visits between labs. These scientists tend to publish the majority of their publications together, while other scientists have more ad-hoc collaborative relationships with OECD based scientists. It is the former category of scientists – which encompasses 62% of the study sample – who have embedded relationships with OECD based scientists that I consider prior OECD connections. That being said, variations of the definition of 'connected to OECD' are used – in particular defining this group as those with more than 50%, 60% and 90% of their publications with OECD coauthors – with qualitatively similar results.

more administrative. He called on us. The next thing we knew we were doing in-country training at his behest..... He was able to nurture a mini [American institution] back home. He was able to do that because of his position.'

Results in Table 8 are consistent with the qualitative evidence. For a sub-sample of returnees for whom full career information was obtained, cross-tabs of the post-pre difference in treated non-migrant outcomes are calculated according to the role that the returnee assumes on returning home. Table 8 shows that there is a larger positive change for non-migrants who have returnees who are occupying administrative positions on their return home. Although the sample of returnees is extremely small, and thus any findings must be taken with a grain of salt, this suggests that those in an administrative position are able to exert greater positive spillovers onto the non-migrants in their institution.

Alternate Explanations

I consider two alternative explanations that could be driving the observed effect. Namely, team work with the returnee, and knowledge flows from the returnee.

Team Work Benefits

Science is increasingly carried out in teams (Wuchty et al 2007). Prior work finds that the co-location of scientists results in increased collaboration rates, and more correlated research trajectories (Catalini 2018). On this basis, and the significant frictions associated with collaborating with scientists in developing countries, the return home of a scientist should result in increasing rates of collaboration between the returnee and the non-migrants. This increased collaboration could increase the rate of publication outcomes, particularly of those in the field of the trainee, due to improvement in the skills within the team. However, theory relating to complementary skills suggests that the formation of teams between a specialist trained abroad, and a generalist trained in the home country may be challenging (Jones 2008). Furthermore - assortative matching theory suggests that the combined output is that of the least productive member of the team (Jones et al 2008; Ahmadpoor and Jones 2018), dis-incentivising the returnee to collaborate with home country scientists. Consistent with Catalini 2018, Figure 10 illustrates that treated scientists are much more likely to co-author with the returnee following their return. However, very few people actually collaborate with the returning trainee, just 66 treated scientists (less than 4%). And this is mostly people who collaborated with FIC AITRP trainees before the event. I consider if it is these people driving the result in Table 9. Splitting the sample into two groups: those who have pre-return characteristics that are correlated with collaborating with the returnee (columns 1 and 3) and those who don't (columns 2 and 4). If benefits from team work with the returnee are driving the main result, I would expect that those with the characteristics correlated with collaborating with the returnee are also the ones who benefit the most from the arrival of the returnee in their institution. However, table 9 shows a different story. Those scientists who are less likely to co-author with the returnee (or counterfactual) experience the greatest positive impact from the return event. Furthermore, the results are robust to removing publications coauthored with the returnee (Table 4, column 2). I therefore do not think that team work benefits are driving the observed results.

Knowledge Flows from the Returnee

Economic geography has long documented a relationship between physical proximity and knowledge transfer (Jaffe et al 1993). Mobile scientists carry knowledge with them, and knowledge flows in the form of citations to a mover's pre-move publications are found to increase in the destination following a move (Azoulay et al 2012; Ganguli 2015). If this is new knowledge, this could improve publication outcomes of non-migrants. I measure changes in citation rates to the returning trainee's (or counterfactual) pre-graduation publications. If benefits from knowing about the returnee's research are driving the main result, this may cast doubt on the hypothesis that non-migrant benefit from the formation of a core/periphery bridge, as it could just be due to new knowledge coming into the institution, irrespective of the involvement of the core. However, table 5 column 6 illustrates the citation rates to the returnee's publications do not increase for treated scientists following the return.

Robustness and Sensitivity Checks

The main threat to identification in this study is the possibility that the treated institutions are getting better, and particularly that they are becoming more internationally connected, just prior or at the same time as the return home of the Fogarty trainees. A few tests help to understand if this is driving the observed results. First, I re-match the 1,740 treated scientists with the same individual level pre-return covariates, and this time overlay pre-return institution level covariates of institution size, productivity and collaboration rates with OECD country institutions. This results in a smaller sample of treated and control scientists (due to the difficulty in finding a similar scientist in a similar institution at the same time) of 2,780 scientists (matches are found for 80% of the treated scientists). I run the main specification regressions on this sample of individual and institution-level matched scientists in Table 10. The results are very similar on this matched sample, providing support that the effect is not driven by selection of the returnees to better performing, or better connected, institutions.

Table 11 provides additional evidence to ascertain the robustness of the results. First I verify that the effect is not driven by a few returnees affecting a large number of non-migrants. Column (2) provides the estimate without returnees who affect large numbers of non-migrants. I remove scientists who have a returnee (and counterfactual) that is in the 95th percentile in terms of the number of non-migrants they impact. The finding is robust and actually greater without these returnees. To provide further evidence that the effect is driven by the returnee and not other institutional factors, I verify that the effect is sensitive to the qualification of the returnee, by removing those returnees who received a Ph.D during their studies in the United States in column (3). As expected, the coefficient is smaller than the baseline which is in line with expectations that those returnees with more experience in the United States, and thus more embedded in the network, exert a greater spillover. Column (4) includes country time trends, and column (5) institution time trends, to remove concerns that the effect is driven by improvements in country level, or institution level capacity that coincides with the timing of the return. The inclusion of the time trends doesn't change the coefficient of interest by much, which is reassuring, but it does increase the standard errors (which is not surprising as it is a demanding specification) which leaves an insignificant finding. To verify that the control sample is not contaminated by the treatment as well — biasing the result, column (6) removes from the sample those scientists that are ever collaborators with the treated scientists. The large increase in the coefficient following the removal of this group of contaminated control scientists suggests that their inclusion biases the result downwards, and so the main result is a conservative estimate. Finally, in columns (7) and (8), I conduct simulation studies to validate the quasi-experiment exploited in the paper. In column (7) I keep just the pre-event data, and generate a placebo return year two years prior to the actual return year. I run the baseline specification with this placebo return year and find a precisely estimated zero effect. This is reassuring that the returnees are not returning to institutions that are improving in the years before the return. In column (8), I keep the control sample only and generate placebo return years for control scientists, where dates are drawn at random from the empirical distribution of return events for the actual returnees. I replicate the main specification but limit the sample to the set of 1,657 control scientists. The effect of return is again reassuringly precisely estimated at zero.

Attrition of trainees arising from the use of just trainees with a publication record post-graduation creates two potential concerns. First, the results could suffer from selection bias. Trainees without publication records following program participation, or those not returning home, differ systematically from those that return home and continue to publish. Although the data on pre-return characteristics are limited, trainees who return and publish are more likely to have studied for a PhD in the United States and to have published prior to their graduation. They are also more likely to be from a country with a greater level of scientific capacity. There are no significant differences in the period of the fellowship or the US institution that they attended. If these differences are indicative of differences in a trainee's potential impact on non-migrants, the results would be biased and should be interpreted as conditional on trainees returning home and continuing to actively publish. I run regressions with interactions of pre-return trainee covariates to assess the hypothesis that pre-return characteristics of trainees affects the magnitude of the impact on non-migrants (Appendix B: Table B1). There is no discernible difference in impact according to the returnee pre-return characteristics. One point to note, however, is that although the main result is relatively stable to inclusion of covariates of the returnee publication record during their fellowship and PhD degree status, the inclusion of a dummy variable reflecting scientific advancement of the home country reduces the main coefficient. Although inconclusive due to the noisy nature of the estimates, this does suggest that the findings in this paper are more relevant for low income countries with relatively more advanced scientific capacity, which is an interesting avenue for future research. Second, those without a publication record may move to institutions in which non-migrants in the control group are working. This is a threat to identification because the control group may be affected by the treatment, although the implications depend on how they impact non-migrants. Unfortunately, this is not testable, but a lack of research productivity, and alumni surveys finding that many take up senior positions in government or non-profits (which are not the same institutions as the control group) suggests that their impact may be minimal.

3.6 Discussion

This paper offers a new perspective on the consequences of return migration of high-skilled workers from developed to developing countries, exploring how returnees can provide a bridge in the network affecting the performance of developing country non-migrants. Through an examination of the impact of the return home of African scientists after training in the United States on non-migrant scientists working in the institutions they return to, the results show that the publication rates of African non-migrant scientists increase following the return event. Furthermore, this increase in publications is mostly in the field of study of the returning scientist. The relationship is contingent on a lack of prior connectivity of the non-migrant.

These findings shed light on the phenomenon of association with a broker, a potentially critical but under-recognized mechanism that shapes the performance of outsiders. Although extant research has long explored the impact of networks on performance, particularly in knowledge production, it has generally considered actors within a network as a function of their ties. Failure to account for the notion that broader features of a network — including the structure (in particular where there are insiders and outsiders - or core/periphery), as well as indirect ties — can also affect performance would lead to an incomplete understanding of how individuals affect the performance of others.

To examine the effects of association with a broker within a core/periphery network, I introduce the concept of a core/periphery bridge. With many network exhibiting a core/periphery structure, the impact of an individual bridging the core and periphery is less well understood. Beyond documenting the performance implications of association with a core/periphery bridge for periphery actors, I also provide supporting evidence of the existence of two concurrent mechanisms driving the impact: (1) accessing the broker's knowledge, and (2) accessing the connections of the broker.

The findings inform the topical debate on immigration and mobility, in particular the impact of migration of high-skilled individuals in and out of developing countries. To date, much research on this topic has focused on the mobile individual themselves. The clarification of conditions under which non-migrants benefit from return migration of others, and when they can access the knowledge and connections of a return migrant — or broker — provides an opportunity to analyse spillovers from migration. The extent to which second-hand brokers can adopt their own brokerage position in the absence of the return migrant, the codified nature of knowledge being transferred, as well as the motives and incentives of the return migrant to both share knowledge and connections should be important considerations of future research and program design aimed to promote sharing of a returnee's access.

The magnitude of the results, as well as the finding that the impact is greatest for more peripheral scientists, are consistent with research on the removal of frictions to accessing inputs for scientific production. Ding et al (2010), for example, studied the impact of the arrival of information technology, in the form of BITNET and Domain Name System, in an American institution on scientists' productivity and collaboration patterns. They found that the arrival of the internet had a positive impact on publication and collaboration rates, particularly for female or scientists in lower tier institutions. That improvements in scientists' access to inputs particularly affects those traditionally more marginalized suggests a need for policy interventions targeting access to collaborations, resources and knowledge for 'outsiders' in an innovation system.

This study has four major limitations. The main challenge, as with many network studies is assessing whether the shock affected the control scientists as well. Given that the control and treated scientists are loosely in the same network (global science) it is plausible to think that there may be ripple effects. If it affects them positively — I provide an underestimate of the main effects. Limitations in the data also prevent me from controlling for the return home of non-FIC AITRP scientists. If other returnees are arriving in the institutions at the same time this could be problematic for the estimate. Second, the FIC AITRP studied is specifically designed to engage developing country sites. Thus the question of whether these results are generalizable to other forms of training programs is unclear. Third, the study is limited to returning HIV researchers in Africa in a time when HIV research, particularly that done in or on African populations, was very topical. Whether these findings are generalizable to other fields and countries is also unclear. Fourth, I use publication records as a proxy for performance. Whether this is a true reflection of performance is unclear. It is difficult to distinguish between capacity to do science, and a capacity to publish in international journals. Returnees and international collaborators could be sharing knowledge that is more centered around the publication process (as opposed to scientific knowledge), which could explain the rise in publication output. Moreover, particularly concerning is the fact that first (and last) authored publications do not significantly increase following return events. This raises questions on whether capacity is actually improving — or whether the relationship with American-based scientists is one in which African scientists carry out low-skilled field work tasks to deliver samples to the United States for analysis. Future work using more reflective measures of actual scientific capacity is necessary.

Despite the above mentioned limitations, the findings have important implications for the valuation of international training programs and policies on bringing people home, decisions on engagement with international scientists and for developing country science more generally. The majority of impact evaluations and cost benefit calculations from international training programs of developing country high-skilled workers just pertain to the individual trained. The benefits to non-migrants estimated in this study suggest that re-conceptualizing the unit of analysis of such estimates is important to understand 'bang for the buck' of programs sending individuals abroad, and the relative merits of incentivizing the return home of high-skilled nationals.

A back of the envelope calculation finds that one returnee contributes around 4 additional publications of non-migrants in the five years after returning. With the returnee publishing an average of 5 publications in the five years after return, this spillover effect is 80% of the effect of the returnee themselves on the institution's innovation output. Given that the average cost of one within sample Fogarty trainee is around USD $144,000 (in today's USD), this suggests that the cost of one African publication, including the returnee's publications as well as the spillover, is approximately USD $16,000. While this estimate does not incorporate re-direction of research funding dollars to the treated institutions, and so is not an absolute measure of cost per African publication, it is interesting to note that this is a fraction of estimates of the cost of other programs designed to encourage publication output. As two examples, Myers (2018) estimates that the cost to the NIH of a single publication coming out of their R01 grant program is between USD $344,000 and $665,000, and Jacob and Lefgren (2011) estimate the cost to the NIH of one publication is $1.7 million. It is important to note as well that my estimate is likely to be an underestimate of the spillover effect. The sample of scientists treated by the returnee in this study is very narrowly defined. I only consider those already publishing at the time of the returnee, thus this estimate doesn't include future students of the returning trainee.

Not only are there a plethora of programs, initiatives and strategies around the world targeting global training or experience for individuals, employees and firms, but the findings of this study are general and applicable to a range of settings. First, I anticipate that alternative programs and policies creating core/periphery bridges under similar conditions would result in similar outcomes. Second, the findings are relevant across a variety of settings that exhibit core/periphery structures. As just one example, hiring strategies of firms, particularly of entrepreneurial firms, should consider how to best leverage bridges with the central network.

This paper raises more questions than it answers. Is the same phenomenon observed if the foreign trained remains abroad? A body of research on the benefits of a skilled diaspora to developing countries has documented the role these individuals play in knowledge flows, remittances and trade. But an understanding of the dynamics in terms of the network is less well understood. Are bridges to the center more important for periphery actors operating in industries or settings that have a dense center? And what happens in settings with more tacit knowledge, or uncertainty? Future work should seek to explore these questions and further our understanding through leveraging the concept of a core/periphery bridge.

References

Agrawal, A., Cockburn, I., McHale, J., 2006 *Gone But Not Forgotten: Labor Flows, Knowledge Network Benefits and Enduring Social Capital* Journal of Economic Geography 6(5): 571-591

Agrawal, A., et al., 2011 *Brain Drain or Brain Bank? The Impact of Skilled Emigration on Poor Country Innovation* Journal of Urban Economics 69 1:4355

Ahmadpoor, M., Jones, B.F., 2018 *Team Output and Individual Productivity in Science and Invention* Working Paper

Azoulay, P., Zivin, J.G., Wang, J., 2010 *Superstar Extinction* Quarterly Journal of Economics 125(2): 549-589

Azoulay, P., Graff Zivin, J.S., Sampat, B.N., 2012 *The Diffusion of Scientific Knowledge Across Time and Space: Evidence from Professional Transitions for the Superstars of Medicine* In The rate and direction of inventive activity revisited, ed. Josh Lerner and Scott Stern, 107-55. Chicago: University of Chicago Press

Azoulay, P., Stuart, T., Wang, Y., 2013 *Matthew: Effect or fable?* Management Science 60(1): 92-109

Blau, P.M., 1964 *Exchange and Power in Social Life* New York: Wiley

Borgatti, S.P., Everett, M.G. 2000 *Models of Core/Periphery Structures* Social Networks 21(4): 375-395

Borjas, G.J., Bratsberg, B., 1996 *Who Leaves? The Outmigration of the Foreign-Born* Review of Economics and Statistics 78: 165-176

Borjas, G.J., Doran, K.B., 2012 *The Collapse of the Soviet Union and the Productivity of American Mathematicians* The Quarterly Journal of Economics 127(3): 1143-1203

Boudreau, K.J., Guinan, E.C., Lakhani, K.R., Riedl, C., 2016 *Looking Across and Looking Beyond the Knowledge Frontier: Intellectual Distance, Novelty, and Resource Allocation in Science* Management Science 62(10): 2765-2783

Boudreau, K.J., et al. 2017 *A Field Experiment on Search Costs and the Formation of Scientific Collaborations* Review of Economics and Statistics 99(4): 565-576

Burt, R.S., 1997 *The Contingent Value of Social Capital* Administrative Science Quarterly 339-365

Burt, R.S., 1998 *The Gender of Social Capital* Rationality and Society 10(1): 5-46

Burt, R.S., 2000 *The Network Structure of Social Capital* Research in Organizational Behavior 22: 345-423

Burt, R.S. 2002 *Bridge Decay* Social Networks 24(4): 333-363

Burt, R.S., 2007 *Secondhand Brokerage: Evidence on the Importance of Local Structure for Managers, Bankers, and Analysts* Academy of Management Journal 50(1): 119-148

Burt, R.S., 2010 *Neighbor Networks: Competitive Advantage Local and Personal* Oxford University Press

Catalini, C., 2018 *Microgeography and the Direction of Inventive Activity* Management Science 64(9): 4348-4364

Cattani, G., Ferriani, S., 2008 *A Core/Periphery Perspective on Individual Creative Performance: Social networks and cinematic achievements in the Hollywood film industry* Organization science 19(6): 824-844

Cattani, G., Ferriani, S., Allison, P.D., 2014 *Insiders, Outsiders, and the Struggle for Consecration in Cultural Fields: A core-periphery perspective* American Sociological Review 79(2): 258-281

Choudhury, P., 2015 *Return Migration and Geography of Innovation in MNEs: A Natural Experiment of Knowledge Production by Local Workers Reporting to Return Migrants* Journal of Economic Geography 16: 585-610

Clauset, A., Arbesman, S., Larremore, D.B., 2015 *Systematic Inequality and Hierarchy in Faculty Hiring Networks* Science advances 1(1): e1400005.

Cole, J.R., Cole, S., 1973 *Social Stratification in Science* Chicago: University of Chicago Press

Crane, D., 1965 *Scientists at Major and Minor Universities: A Study of Productivity and Recognition* American Sociological Review 30(5): 699-714

Dasgupta, P., David, P.A., 1987 *Information Disclosure and the Economics of Science and Technology* Arrow and the ascent of modern economic theory. Palgrave Macmillan, London, 519-542

Ding et al., 2010 *The Impact of Information Technology on Academic Scientists' Productivity and Collaboration Patterns* Management Science 56(9): 1439-1461

Dustmann, C., Fadlon, I., Weiss, Y., 2011 *Return Migration, Human Capital Accumulation and the Brain Drain* Journal of Development Economics 95(1): 58-67

Fafchamps, M., Leij, M.J., Goyal, S., 2006 *Scientific Networks and Co-authorship* Working Paper

Fry, C.V., 2019 *'Never Let a Good Crisis go to Waste': The 2014 Ebola Epidemic and the Careers of West African Scientists* Working Paper

Ganguli, I., 2015 *Immigration and Ideas: What Did Russian Scientists "Bring" to the United States?* Journal of Labor Economics 33.S1: S257-S288

Gaulé, P., 2014 *Who Comes Back and When? Return Migration Decisions of Academic Scientists* Economics Letters 124(3): 461-464

Giannetti, M., Liao, G., Yu, X., 2015 *The Brain Gain of Corporate Boards: A Natural Experiment from China* The Journal of Finance 70(4): 1629-1682

Gibson, J., McKenzie, D., 2014 *Scientific Mobility and Knowledge Networks in High Emigration Countries: Evidence from the Pacific* Research Policy 43(9): 1486-95

Granovetter, M.S., 1973 *The Strength of Weak Ties* The American Journal of Sociology 78(6): 1360-1380

Goldsmith-Pinkham, P., Imbens, G.W., 2013 *Social Networks and the Identification of Peer Effects* Journal of Business & Economic Statistics 31(3): 253-264

Iacus, S.M., King, G., Porro, G., 2011 *Multivariate Matching Methods that are Monotonic Imbalance Bounding* Journal of the American Statistical Association 106(493): 345-361

Hwang, K., 2008 *International Collaboration in Multilayered Center-Periphery in the Globalization of Science and Technology* Science, Technology and Human Values 33(1): 101-133

Hunt, J., Gauthier-Loiselle, M., 2010 *How Much Does Immigration Boost Innovation?* American Economic Journal: Macroeconomics 2(2): 31-56

Jackson, M.O., Wolinsky. A., 1996 *A Strategic Model of Social and Economic Networks* Journal of Economic Theory 71(1): 44-74

Jacob, B.A., Lefgren, L., 2011 *The Impact of Research Grant Funding on Scientific Productivity* Journal of Public Economics 95(9-10): 1168-1177

Jaffe, A.B., Trajtenberg, M., Henderson, R., 1993 *Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations* The Quarterly Journal of Economics 108: 577-598

Jaffe, A., Trajtenberg, M., 1999 *International Knowledge Flows: Evidence from Patent Citations* Economics of Innovation and New Technology 8: 105-136

Jones, B.F., 2008 *The Knowledge Trap: Human Capital and Development Reconsidered* Working Paper

Jonkers, K., Tjissen, R., 2008 *Chinese Researchers Returning Home: Impacts of International Mobility on Research Collaboration and Scientific Productivity* Scientometrics 77: 309 - 333

Kahn, S., MacGarvie, M.J., 2016a *How Important is U.S. Location for Research in Science?* Review of Economics and Statistics 98(2): 397-414

Kahn, S., MacGarvie, M., 2016b *Do Return Requirements Increase International Knowledge Diffusion? Evidence from the Fulbright Program* Research Policy 45(6): 1304-1322

Kerr, W.R. 2008 *Ethnic Scientific Communities and International Technology Diffusion* The Review of Economics and Statistics 90(3): 518-537

Kogut, B., Macpherson, J.M., 2011 *The Mobility of Economists and the Diffusion of Policy Ideas: The Influence of Economics on National Policies* Research Policy 40(10): 1307-1320

Kolympiris, C., Hoenen, S., Klein, P.G., 2018 *Learning by Seconding: Evidence from NSF Rotators* Working Paper

Latour, B., 1987 *Science in Action* Cambridge, MA: Harvard University Press

Leydesdorff, L., Wagner, C.S., 2008 *International Collaboration in Science and the Formation of a Core Group* Journal of Informetrics 2(4): 317-325.

Levin, S.G., Stephan, P.E., 1991 *Research Productivity over the Life Cycle: Evidence for Academic Scientists* American Economic Review 81: 114-132

Long, J.S., Allison, P.D., McGinnis, R., 1979 *Entrance into the Academic Career* American Sociological Review 44(5): 816-830

Manski, C.F., 1993 *Identification of Endogenous Social Effects: The Reflection Problem* The Review of Economic Studies 60(3): 531-542

Merton, R.K., 1972 *Insiders and Outsiders: A Chapter in the Sociology of Knowledge* American Journal of Sociology 78(1): 9-47

Moser, P., Voena, A., Waldinger, F., 2014 *German Jewish Immigration and US Invention* American Economic Review 104(10): 3222-55

Mohnen, M., 2016 *Stars and Brokers: Peer Effects Among Medical Scientists* Working Paper

Myers, K., 2018 *Managing the Rate and Redirection of Science: Evidence from NIH Data* Working Paper

Obstfeld, D., 2005 *Social Networks, The Tertius Iungens Orientation, And Involvement in Innovation* Administrative science quarterly 50(1): 100-130

Obukhova, E., 2012 *Experience vs. Legitimacy: Returnees and Political Ties in China* Working Paper, Social Science Research Network

Oettl, A., Agrawal, A., 2008 *International Labor Mobility and Knowledge Flow Externalities* Journal of International Business Studies 39(8): 1242-1260

Reagans, R.E., Zuckerman, E.W., 2008 *A Shortcut to Efficiency? Implications of the Small but Stratified World* Working Paper

Rose, M.E. Kitchin, J.R., 2019 *Scopus: Scriptable Bibliometrics using a Python Interface to Scopus* SoftwareX 10: 100263

Saxenian, A., 2006 *The New Argonauts: Regional Advantage in a Global Economy* Cambridge, MA: Harvard University Press

Scellato, G., Franzoni, C., Stephan, P., 2015 *Migrant Scientists and International Networks* Research Policy 44(1): 108-120

Singh, J., 2005 *Collaborative Networks as Determinants of Knowledge Diffusion Patterns* Management Science 51(5): 756-770

Stephan, PE. 2012 *How Economics Shapes Science* Vol. 1. Cambridge, MA: Harvard University Press

Stuart, T.E., Hoang, H., Hybels, R.C., 1999 *Interorganizational Endorsements and the Performance of Entrepreneurial Ventures* Administrative Science Quarterly 44(2): 315-349

Wagner, C.S., Leydesdorff, L., 2005 *Network Structure, Self-organization, and the Growth of International Collaboration in Science* Research Policy 34(10): 1608-1618

Wang, D., 2015 *Activating Cross-Border Brokerage: Interorganizational Knowledge Transfer Through Skilled Return Migration* Administrative Science Quarterly 30(1): 133-176

Watts, D.J., Strogatz, S.H., 1998 *Collective Dynamics of 'Small-world' Networks* Nature 393.6684: 440

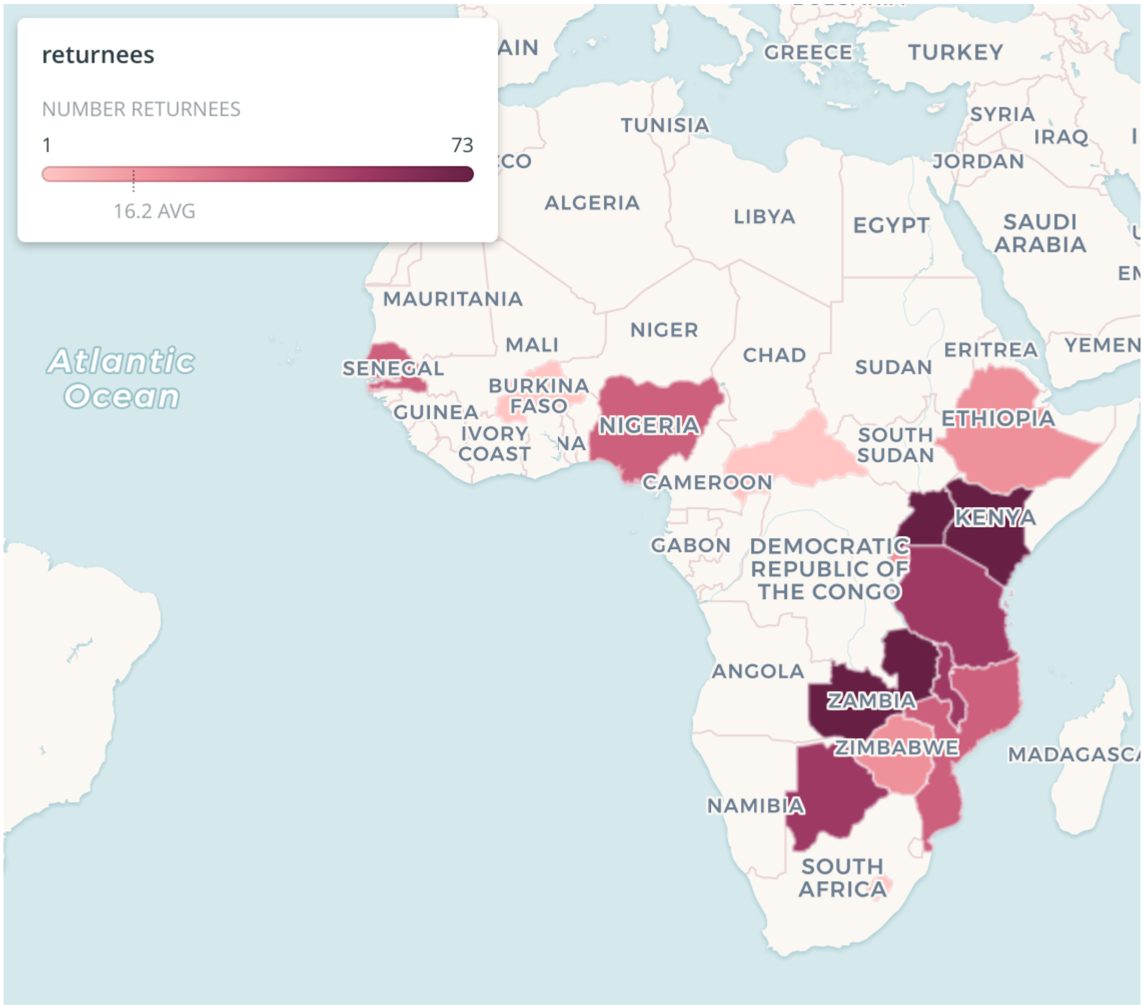
Wuchty, S., Jones, B.F., Uzzi, B., 2007 *The Increasing Dominance of Teams in Production of Knowledge* Science 316(5827): 1036-1039

Zelnio, R,. 2012 *Identifying the Global Core-Periphery Structure of Science* Scientometrics 91(2): 601-615

Zucker, L.G., Darby, M.R., 2007 *Star Scientists, Innovation and Regional and National Immigration* No. w13547. National Bureau of Economic Research

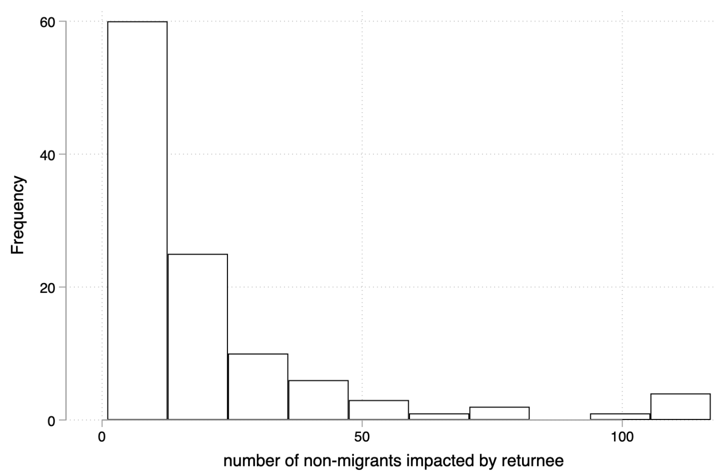
Zuckerman, H., 1988 *The Sociology of Science* In Handbook of Sociology, edited by N.J. Smelser. Newbury Park, CA: Sage

Figure 3: Number of FIC AITRP Trainee Returnees to African Countries



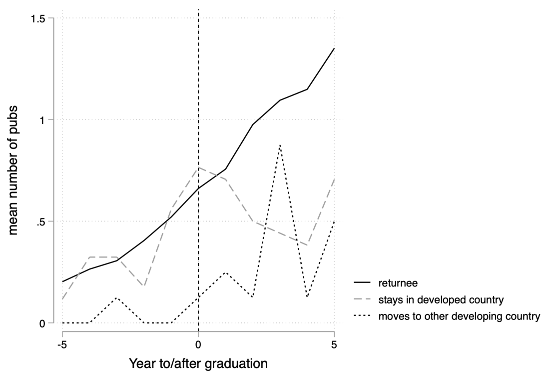
Note: The overall count in each country of the 242 sample FIC AITRP trainees known to return back to Africa between 1988-2014.

Figure 4: Histogram of Number of Non-Migrants Each Returnee Impacts

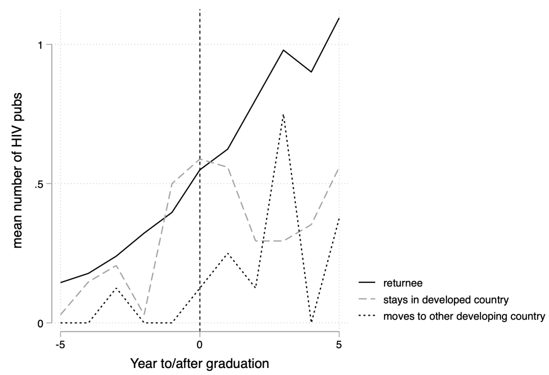


Note: I compute the number of treated non-migrants in the sample affected by each FIC AITRP trainee.

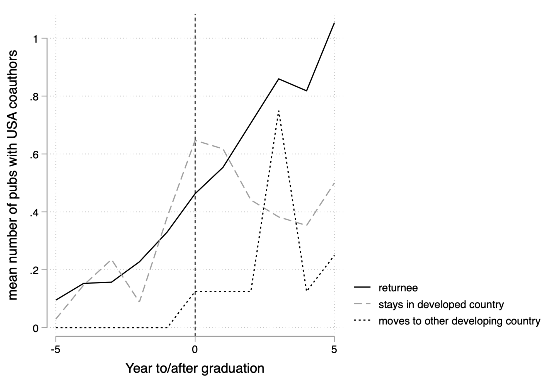
Figure 5: Publication Trends for FIC AITRP trainees



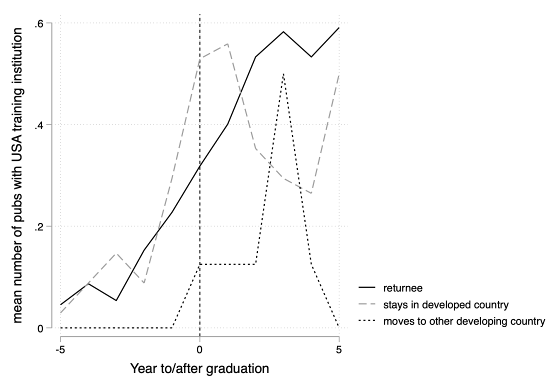
(a) Mean number of publications



(b) Mean number of HIV publications



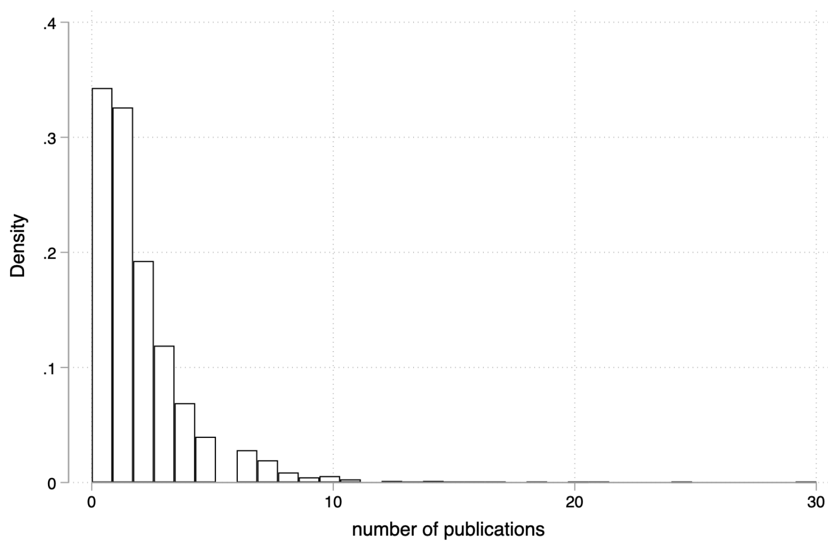
(c) Mean number of publications with American-based coauthors



(d) Mean number of publications with American training institution

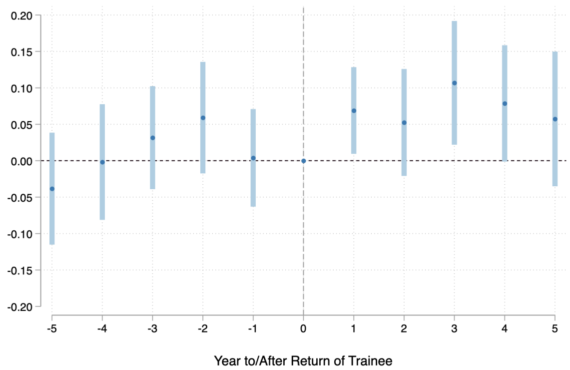
Note: Publication trends for the 284 FIC AITRP trainees who have a publication record following graduation (242 who return home, 34 who remain in a developed country, 8 who move to another African country) are plotted for the five years before and after graduation.

Figure 6: Histogram of Non-Migrant Publication Rate at the Time of Return



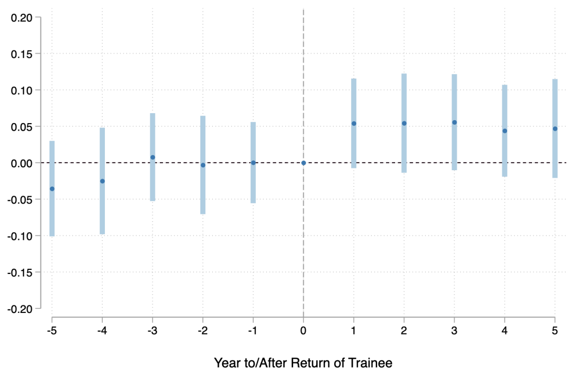
Note: I compute the number of publications in the year of return authored by the 3,314 sample treated and control non-migrant scientists.

Figure 7: Impact of FIC AITRP Trainee Return on Non-Migrant African Scientists' Publication Outcomes

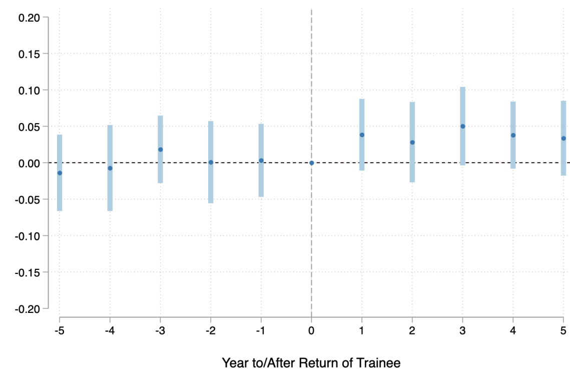


(a) Number of publications

(b) Any publication



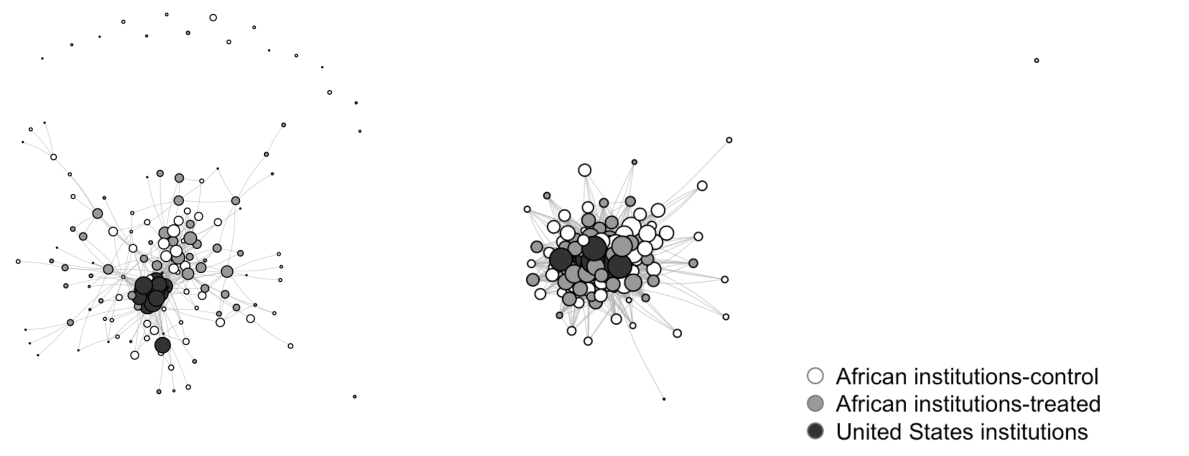
(c) Number of HIV publications



(d) Any HIV publication

Note: The solid blue dots in the above plots correspond to coefficient estimates in panels (a) and (c) stemming from conditional (scientist) fixed effects ordinary least squares specifications in which inverse hyperbolic sine outcomes are regressed onto year effects, article age effects, as well as 10 interaction terms between treatment status and the number of years before/after the return of a trainee (the indicator variable for treatment status interacted with the year of return is omitted). And coefficient estimates stemming from conditional (scientist) fixed effects Linear Probability Model specifications in panel (b) and (d) in which publication dummy variables are regressed onto year effects, article age effects, as well as 10 interaction terms between treatment status and the number of years before/after the return of a trainee (the indicator variable for treatment status interacted with the year of return is omitted). All specifications also include a full set of lead and lag terms common to both the treated and control articles to fully account for transitory trends in citations around the time of the return. The 95% confidence interval robust standard errors clustered around the institution is plotted with light blue bars.

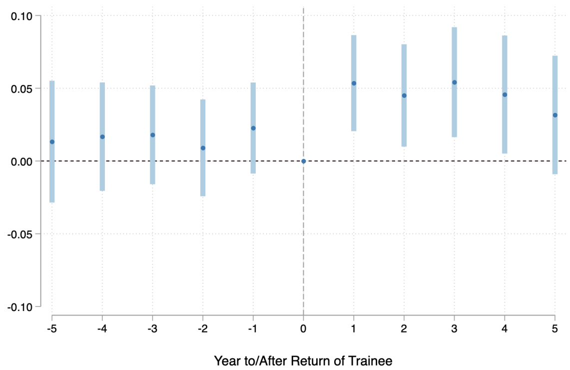
Figure 8: Pre- and Post-Program Collaboration Network of Within Sample African Institutions and American Institutions Involved in FIC AITRP



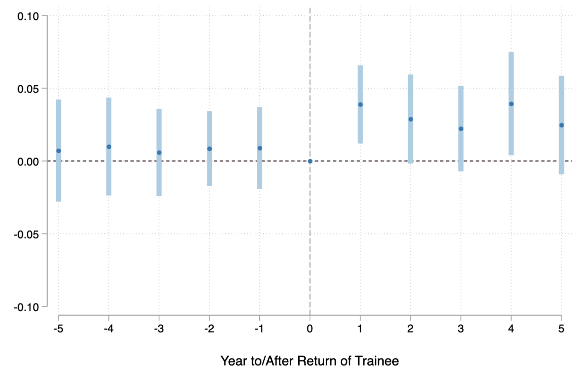
(a) Pre-Program Network (1985-1989) (b) Post-Program Network (2015-2019)

Note: The collaboration network of the publications in medical related subjects of 152 African institutions that are included in the sample (and publish in 1985-1989), and the 14 American institutions involved in the FIC AITRP program (excluding publications authored by FIC AITRP African trainees) is plotted for the pre-program period (1985-1988) in panel (a) and post-program period (2014-2019) in panel (b). American institutions are represented by gray circles, African institutions that are treated between 1988 and 2014 are gray circles, and African institutions never treated between 1988 and 2014 are white circles. The lines between circles represent a collaborative link. The size of the circles in the network are adjusted according to the log of the number of publications they produce in the time frame.

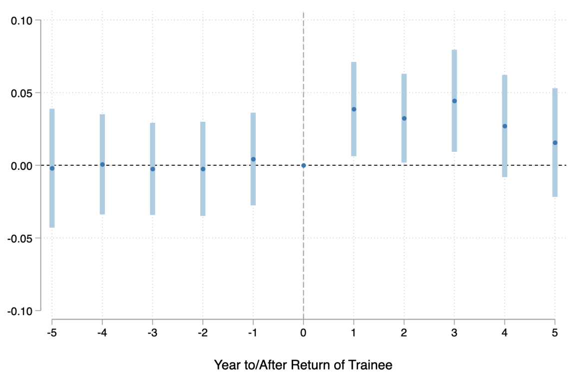
Figure 9: Impact of FIC AITRP Trainee Return on Non-Migrant African Scientists' Collaborations



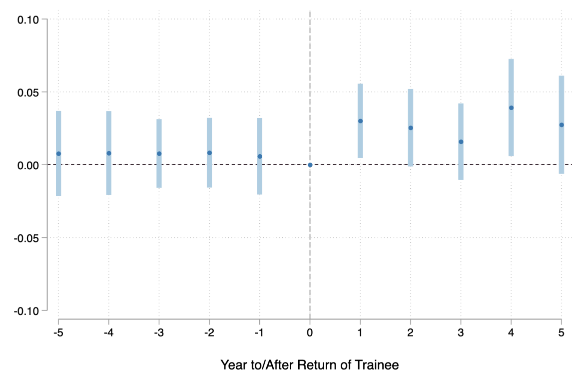
(a) Any publication with United States based coauthors



(b) Any publication with American FIC AITRP institution coauthors



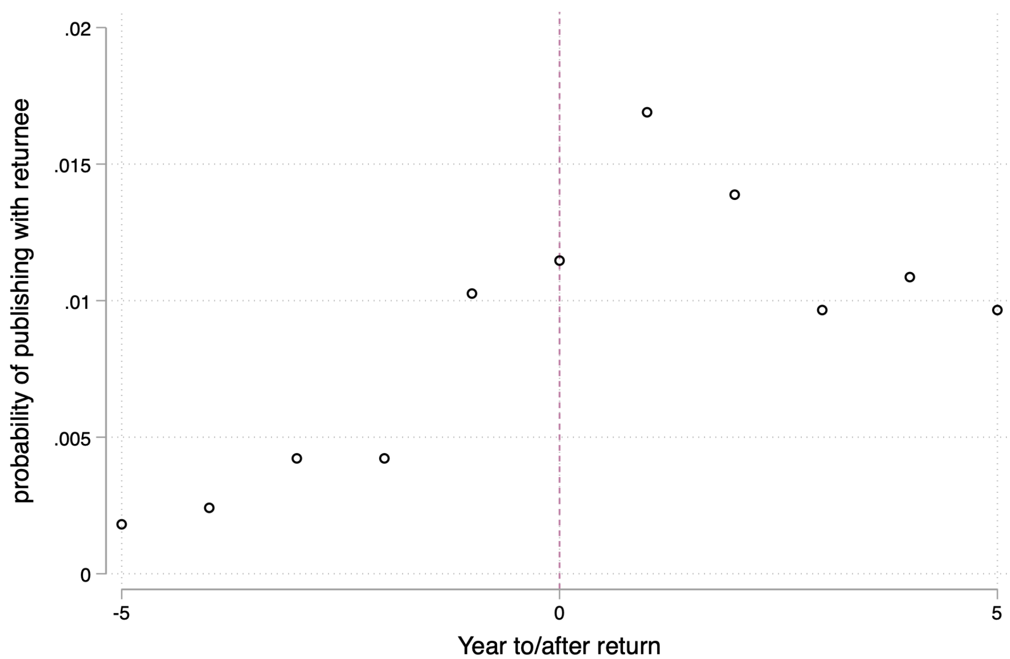
(c) Any new United States based coauthor



(d) Any new American FIC AITRP institution coauthor

Note: The solid blue dots in the above plots correspond to coefficient estimates stemming from conditional (scientist) fixed effects Linear Probability Model specifications in which coauthoring dummy variables are regressed onto year effects, article age effects, as well as 10 interaction terms between treatment status and the number of years before/after the return of a trainee (the indicator variable for treatment status interacted with the year of return is omitted). The specifications also include a full set of lead and lag terms common to both the treated and control articles to fully account for transitory trends in citations around the time of the return. The 95% confidence interval robust standard errors, clustered around the institution is plotted with light blue bars.

Figure 10: Non-Migrant African Scientist's Probability of Publishing with FIC AITRP Trainee Before and After Return Event



Note: The average probability of a scientist in the treated sample to publish with the returning FIC AITRP trainee in each year before and after the return is plotted.

Table 1: FIC AITRP Returnees by Country and American Training Institution

Country	Number of Returnees
Kenya	73
Uganda	60
Zambia	35
Tanzania	20
Botswana	19
Malawi	9
Senegal	6
Nigeria	5
Mozambique	4
Rwanda	3
Zimbabwe	3
Ethiopia	2
Central African Republic	1
Burkina Faso	1
Lesotho	1
American Training Institution	
University of Washington	84
Case Western Reserve University	40
Harvard School of Public Health	26
Johns Hopkins University	21
Vanderbilt University	16
University of Alabama at Birmingham	15
Baylor College of Medicine	8
Dartmouth College	7
Duke University	7
Brown University	5
University of Nebraska, Lincoln	4
Emory University	3
State University of New York at Buffalo	3
University of Maryland Baltimore	3

Note: This table provides details on the sample of 242 scientists who are trained in the United States in long-term training programs supported by the FIC AITRP and return home following their graduation (graduating between 1988 and 2014).

Table 2: Summary Statistics for FIC AITRP Trainee Returnees (N=112)

	mean	median	std. dev.	min.	max.
Year of fellowship	2003	2004	6.71	1989	2014
Year of graduation	2004	2006	7.03	1989	2014
Ph.D degree	0.098	0	0.30	0	1
Masters degree	0.51	1	0.50	0	1
Already published before fellowship	0.39	0	0.49	0	1
Career age at fellowship if already published	3.53	2	4.20	0	22
Lag between graduation and publication in home country	3.95	3	3.79	1	22
Return to institution with broad institution program	0.65	1	0.48	0	1
Post graduation number of publications	5.56	3	5.84	0	24
Post graduation number of HIV publications	4.54	2	5.26	0	22
Publish with U.S. coauthors post graduation	0.76	1	0.43	0	1
Post graduation number of publications with U.S. coauthors	4	2	5.25	0	24
Publish with any U.S. training institution coauthors post graduation	0.71	1	0.46	0	1
Post graduation number of publications with any U.S. training institution coauthors	3.48	2	4.79	0	22
Publish with own U.S. training institution coauthors post graduation	0.63	1	0.49	0	1
Post graduation number of publications with own U.S. training institution coauthors	3.02	1	4.67	0	22
Publish with return institution coauthors post graduation	0.75	1	0.43	0	1
Post graduation number of publications with return institution coauthors	3.75	2	4.85	0	23

Note: The sample consists of 112 scientists who are trained in the United States in long-term training programs supported by the FIC AITRP and return home following their graduation (graduating between 1988 and 2014). These scientists are the first to return to an institution during the career of a sample non-migrant scientist. Post graduation publications are those published in the five years following the graduation date.

Table 3: Statistics for Non-Migrant African Study Scientists the Year of the Return of a Trainee

	Control Scientists (N = 1,657)					Treated Scientists (N = 1,657)				
	mean	median	std. dev.	min.	max.	mean	median	std. dev.	min.	max.
Career age	9.98	8	7.51	2	42	9.87	8	7.49	2	42
Number of publications	6.84	5	7.38	0	97	6.56	4	7.33	0	66
Number of last author publications	1.38	0	2.90	0	39	1.16	0	2.59	0	38
Number of journal impact factor (JIF) weighted publications	6.02	3.03	9.43	0	111.56	6.85	3.33	11.48	0	123.75
Number HIV publications	2.54	1	3.67	0	43	2.75	1	4.14	0	56
Number publications with returning trainee	0.0012	0	0.35	0	1	0.013	0	0.11	0	1
Number publications with U.S. coauthors	1.47	0	3.05	0	50	1.80	0	3.89	0	42
Number publications with any U.S. training institution	0.57	0	1.71	0	22	0.81	0	2.67	0	42
Number publications with returnee's U.S. training institution	0.37	0	3.22	0	80	1.74	0	10.93	0	235

Note: This study sample consists of 3,314 African scientists who were actively publishing in HIV related research at the time of the return (or counterfactual return) of a FIC AITRP trainee. All variables are measured using scientist level data gathered from the Elsevier Scopus database, and measurements are made for the five years before the return, unless stated otherwise stated. U.S. training institutions include institutions participating in the FIC AITRP program and accepted long-term trainees between 1988-2014.

Table 4: Impact of FIC AITRP Trainee Return on Non-Migrant African Scientists' Publication Outcomes

	Ordinary least squares model (IHS-transformed dependent variable)				Linear probability model			
	(1) num pubs	(2) num pubs without returnee	(3) JIF weighted pubs	(4) num HIV pubs	(5) any pub	(6) any HIV pub	(7) any first authored pubs	(8) any last authored pubs
AFTER RETURN × RETURN INSTITUTION	0.062*	0.061*	0.062*	0.059**	0.023	0.036**	0.0091	0.013
	(0.034)	(0.033)	(0.030)	(0.022)	(0.018)	(0.016)	(0.010)	(0.011)
Mean of the dependent variable	1.53	1.53	1.67	0.61	0.58	0.32	0.19	0.17
Number of scientists	3,314	3,314	3,314	3,314	3,314	3,314	3,314	3,314
Number of scientists × year observations	36,454	36,454	36,454	36,454	36,454	36,454	36,454	36,454
Number of institutions	440	440	440	440	440	440	440	440

[a] Estimates in columns (1)-(4) stem from fixed effects ordinary least square specifications with dependent variables being inverse hyperbolic sine of counts of outcomes per scientist in a given year. Estimates in columns (5)-(8) stem from fixed effects linear probability model specifications with dependent variables being dummy outcomes of the outcome occurring. All models incorporate a full suite of calendar year, career age and scientist fixed effects.
[b] Heteroskedastic robust standard errors, clustered at the institution, are given in parentheses. *, **, *** denote statistical significance at p-values of 0.1, 0.05 and 0.01 respectively.

Table 5: Knowledge Flows from Returnee's International Network to Non-Migrant African Scientist

	(1) cite any U.S.	(2) cite any U.S. training institution	(3) cite specific U.S. training institution	(4) cite specific U.S. training institution PI	(5) cite specific U.S. training institution (without training institution coauthors)	(6) cite returnee
AFTER RETURN × RETURN INSTITUTION	0.039**	0.034**	0.022**	0.0069*	0.018**	-0.00034
	(0.018)	(0.016)	(0.0087)	(0.0039)	(0.0081)	(0.0019)
Mean of the dependent variable	0.45	0.34	0.075	0.020	0.071	0.0028
Number of scientists	3,314	3,314	3,314	3,314	3,314	3,314
Number of scientists × year observations	36,454	36,454	36,454	36,454	36,454	36,454
Number of institutions	440	440	440	440	440	440

[a] Estimates stem from fixed effects linear probability models with specifications with dependent variables being dummy outcomes. Models incorporate a full suite of calendar year, career age and scientist fixed effects. U.S. training institutions include institutions participating in the FIC AITRP program and accepted long-term trainees between 1988-2014.
[b] Heteroskedastic robust standard errors, clustered at the institution, are given in parentheses. *, **, *** denote statistical significance at p-values of 0.1, 0.05 and 0.01 respectively.

Table 6: Impact of FIC AITRP Trainee Return on Non-Migrant African Scientists' Collaborations

	(1) collaborate with U.S. coauthor	(2) collaborate with any U.S. training institution coauthor	(3) collaborate with specific U.S. training institution coauthor	(4) collaborate with specific U.S. training institution PI	(5) any new U.S. coauthors	(6) any new specific U.S. training institution coauthors
AFTER RETURN × RETURN INSTITUTION	0.032***	0.024**	0.0087**	-0.00012	0.032***	0.0061*
	(0.012)	(0.0097)	(0.0043)	(0.0016)	(0.010)	(0.0035)
Mean of the dependent variable	0.22	0.11	0.026	0.0062	0.20	0.092
Number of scientists	3,314	3,314	3,314	3,314	3,314	3,314
Number of scientists × year observations	36,454	36,454	36,454	36,454	36,454	36,454
Number of institutions	440	440	440	440	440	440

[a] Estimates stem from fixed effects linear probability model specifications with dependent variables being dummy outcomes. All models incorporate a full suite of calendar year, career age and scientist fixed effects. U.S. training institutions include institutions participating in the FIC AITRP program and accepted long-term trainees between 1988-2014.

[b] Heteroskedastic robust standard errors, clustered at the institution, are given in parentheses. *, **, *** denote statistical significance at p-values of 0.1, 0.05 and 0.01 respectively.

Table 7: Breakdown of Publication Outcomes by Non-Migrant African Scientist Network Characteristics

	Ordinary least squares model (IHS-transformed dependent variable)			Linear probability model		
	(1)	(2) num pubs	(3)	(4)	(5) HIV pub	(6)
	connected to training institution	connected to OECD	not connected	connected to training institution	connected to OECD	not connected
AFTER RETURN × RETURN INSTITUTION	0.033	0.0001	0.076*	0.018	0.012	0.041**
	(0.14)	(0.038)	(0.041)	(0.066)	(0.023)	(0.017)
Mean of the dependent variable	2.50	1.62	1.39	0.55	0.38	0.27
Chi^2	0.11, p>.05	2.23, p>.05		0.13, p> .05	1.45, p> .05	
Number of scientists	195	1,128	1,991	195	1,128	1,991
Number of scientists × year observations	2,145	12,408	21,901	2,145	12,408	21,901
Number of institutions	64	246	276	64	246	276
Number of returnees	63	90	103	63	90	103

[a] The sample of non-migrant treated and control scientists is split into three groups: those who have published with the returning trainee's U.S. training institution before the return (or counterfactual) in columns (1) (4), those who have published with OECD collaborators in over 75% of their publications in the 5 years before the return year (or counterfactual) (but never the training institution) - columns (2) (5), and those who have less than 75% of their publications with an OECD collaborator in the 5 years before the return event (or counterfactual) in columns (3) (6). The Chi^2 value represents significance tests between the three models that use distinct samples. I compare models 1 and 2 with model 3. [b] Estimates stem from fixed effects ordinary least square specifications with dependent variables being inverse hyperbolic sine of counts of outcomes per scientist per year in columns (1)-(3), and fixed effects linear probability model specifications with dummy outcomes in columns (4)-(6). All models incorporate a full suite of calendar year, career age and scientist fixed effects.
[c] Heteroskedastic robust standard errors, clustered at the institution, are given in parentheses. *, **, *** denote statistical significance at p-values of 0.1, 0.05 and 0.01 respectively.

Table 8: Difference in Non-Migrant Scientist Change in Publication Rate by Returnee Role

	Returnee Role	
	Administrative	Technical/Teaching
Average difference between post-return and pre-return number of publications (sd)	0.70 (8.10)	-0.63 (8.09)
Number of scientists	30	493
Number returnees	4	29

Note: The sample of treated scientists for whom full resume information on the returnee was found was split into two groups: those who have a returnee taking up an administrative position in column (1) and those who have a returnee taking up a technical or teaching position in column (2). The average difference across scientists in each group between the number of pre and post- return publications is given.

Table 9: Breakdown of Outcomes by Non-Migrant African Scientist Likelihood to Coauthor with Returnee

	Ordinary least squares model (IHS-transformed dependent variable)		Linear probability model	
	(1)	(2)	(3)	(4)
	num pubs		HIV pub	
	high probability	low probability	high probability	low probability
AFTER RETURN × RETURN INSTITUTION	0.015	0.064*	0.024	0.036**
	(0.11)	(0.034)	(0.053)	(0.017)
Mean of the dependent variable	3.06	1.40	0.63	0.30
Chi^2	0.59		0.03	
	p>.05		p>.05	
Number of scientists	261	3,053	261	3,053
Number of scientists × year observations	2,871	33,583	2,871	33,583
Number of institutions	83	431	83	431
Number of returnees	73	112	73	112

[a] A predicted probability of coauthoring with the returning trainee is generated by assigning linear predictions from a fitted logit model of pre-return scientist characteristics (collaboration and publication record) on the probability of collaborating with the returning trainee. The sample of scientists is split into two groups: those who have a high predicted probability (in the 95th percentile of the distribution) to coauthor with the returning trainee (or counterfactual) in columns (1) (3), and the rest of the sample in columns (2) (4).
[b] Estimates stem from fixed effects ordinary least square specifications with dependent variables being inverse hyperbolic sine of counts of outcomes per scientist per year in columns (1) (2). Estimates stem from fixed effects linear probability model specifications with dependent variables being dummy outcomes of the event occurring in columns (3) (4). All models incorporate a full suite of calendar year, career age and scientist fixed effects.
[c] Heteroskedastic robust standard errors, clustered at the institution, are given in parentheses. *, **, *** denote statistical significance at p-values of 0.1, 0.05 and 0.01 respectively.

Table 10: Main Specification Using Reduced Sample with Individual and Institution Re-Return Characteristics Match

	Ordinary least squares model (IHS-transformed dependent variable)	Linear probability model	
	(1) num pubs	(2) any HIV pub	(3) collaborate with any U.S. training institution coauthor
AFTER RETURN × RETURN INSTITUTION	0.062*	0.030	0.018**
	(0.035)	(0.019)	(0.0082)
Mean of the dependent variable	1.50	0.32	0.10
Number of scientists	2,780	2,780	2,780
Number of scientists × year observations	30,580	30,580	30,580
Number of institutions	233	233	233

[a] Estimates in column (1) stem from fixed effects ordinary least square specification with dependent variable being inverse hyperbolic sine of counts of outcomes per scientist in a given year. Estimates in columns (2) and (3) stem from fixed effects linear probability model specifications with dependent variables being dummy outcomes. All models incorporate a full suite of calendar year, career age and scientist fixed effects.
[b] Heteroskedastic robust standard errors, clustered at the institution, are given in parentheses. *, **, *** denote statistical significance at p-values of 0.1, 0.05 and 0.01 respectively.

Table 11: Sensitivity Checks

Dept Var: num pubs	(1) benchmark specification	(2) without 'gregarious' returnees	(3) without Ph.D returnees	(4) with country time trends	(5) with institution time trends	(6) without contaminated controls	(7) placebo test 1	(8) placebo test 2
AFTER RETURN × RETURN INSTITUTION	0.062*	0.079**	0.054*	0.043	0.074	0.19***	-0.011	-0.0088
	(0.034)	(0.031)	(0.032)	(0.034)	(0.045)	(0.035)	(0.026)	(0.033)
Mean of the dependent variable	1.54	1.52	1.54	1.54	1.54	1.40	1.43	1.52
Number of scientists	3,314	2,856	3,314	3,314	3,314	2,419	3,314	1,657
Number of scientists × year observations	36,454	31,416	36,454	36,454	36,454	26,609	19,884	18,227
Number of institutions	440	411	428	440	440	314	440	391

[a] Estimates stem from fixed effects ordinary least square specifications with dependent variables being inverse hyperbolic sine of counts of publications per scientist per year. All models incorporate a full suite of calendar year, career age and scientist fixed effects. [b] Column (1) provides the baseline specification; (2) removes non-migrants who are impacted by returnees (or counterfactual) who are in the 90th percentile in terms of the number of non-migrants they impact; (3) removes returnees with a PhD; (4) includes country time trends; (5) includes institution time trends; (6) removes control scientists who ever coauthor with treated scientists; (7) keeps just pre-event (or counterfactual) data and runs the specification with the same treated scientists with an event date two years prior to the actual event date; (8) keep just control scientists and randomly assigns 'treated' dummy to half of the sample, and treatment dates according to the actual distribution of treatment dates.

[c] Heteroskedastic robust standard errors, clustered at the institution, are given in parentheses. *, **, *** denote statistical significance at p-values of 0.1, 0.05 and 0.01 respectively.

Appendix A: Construction of the Control Group

I detail the procedure implemented to identify the control scientists that help to account for life-cycle and secular trends in the difference-in-differences specification. Publication outcomes might be subject to life-cycle patterns, with outcomes reflecting the trends of the age of the scientist. Also - scientific productivity, particularly in Africa, is rapidly changing over time. Therefore relying on scientists treated earlier or later as an implicit control group may not fully capture these time-varying omitted variables.

To address this concern, I create a sample of control scientists to account for time varying variables in the difference-in-differences specification. Specifically I identify a control scientist who is 'similar' to each treated scientist and assign to them their matched treated scientist's counterfactual return event (returning scientist / return year / American training institution). The control scientists are selected from a universe of possible scientists who are based in FIC AITRP countries and affiliated with institutions that never receive a FIC AITRP returnee in their career lifetime.

The universe of possible control scientists is generated using affiliation data from Elsevier Scopus publication database with inclusion criteria such that the scientist must have published at least three times in their lifetime and at least once as first or last author (to remove technicians or incidental publishers). The institution of each scientist is determined as being the institution in which they are affiliated with in a given time period in over 75% of their publications (to avoid visiting or honorary appointments).

The list of covariates used to identify 'similar' control scientists for each treated scientist such that the following conditions are met:

1. treated scientists exhibit no differential output trends relative to control collaborators up to the time of return (or counterfactual);

2. treated scientists exhibit no differential trends in terms of international, particularly American, collaborations relative to control collaborators up to the time of return (or counterfactual);

3. treated scientists exhibit no differential trends in terms of their field of study relative

to control collaborators up to the time of return (or counterfactual);

4. the distribution of career age at the time of return (or counterfactual) are for similar treated and control scientists.

Coarsened exact matching. To meet these goals, I implement the nonparametric 'coarsened exact matching' (CEM) procedure (Iacus, King and Porro 2011) to identify a control scientist for each treated scientist. The first step is to select a set of covariates on which to guarantee balance, and the second is to create a large number of (coarse) strata that covers the entire support of the joint distribution of the covariates in the previous step. In a third step, each observation is allocated to a stratum and for each treated observation, a control is selected from the same stratum. If there are multiple choices possible, one is selected randomly. If the treated observation is unmatched it is removed from the sample. In this implementation, control scientists are recycled, and so a small number serve as the control for several treated observations (which is accounted for in the specification through the use of scientist specific identifiers by which to cluster standard errors).

Implementation I identify controls based on the following set of covariates (t denotes year of return): career age at t, a dummy for any HIV publication in the four years before t, a dummy for any publications with United States based coauthors in the four years before t, a dummy for any publications with coauthors at American institutions involved in FIC AITRP program in the four years before t, and dummy for any publications with United States based coauthors in years t-1, t-2 ,t-3 and t-4.

I implement the CEM procedure year by year, with replacement. Specifically, in year t, I

1. eliminate from the set of potential controls all those who begin their publication record after year t-1;

2. create the strata using the variables described above;

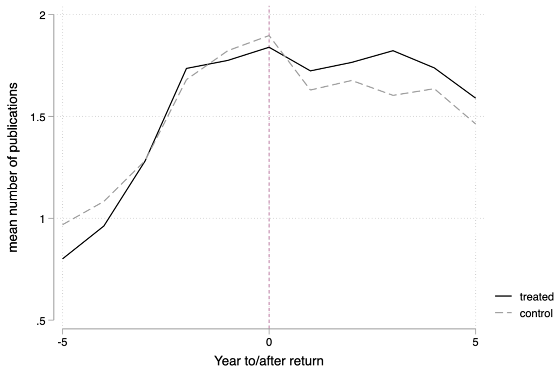
3. identify within each strata a control for each treated unit, randomly selecting one if there are more than one match;

4. assign the control a counterfactual returnee/year of return/returnee American training institution based on the matched treated returnee;

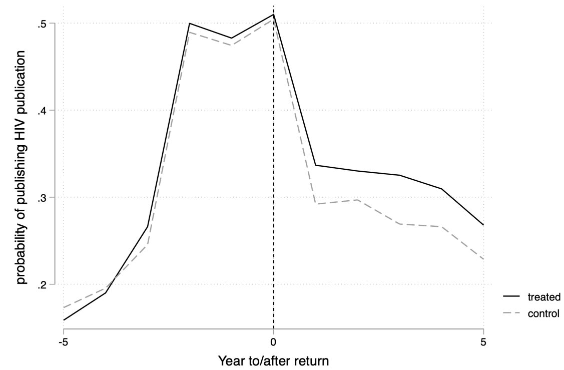
5. repeat these steps for year t + 1

I match 1,657 of 1,740 treated scientists (95%). In the sample of 1,657 treated + 1,657 controls = 3,314 scientists, there is no evidence of preexisting trends in output (figure A.1).

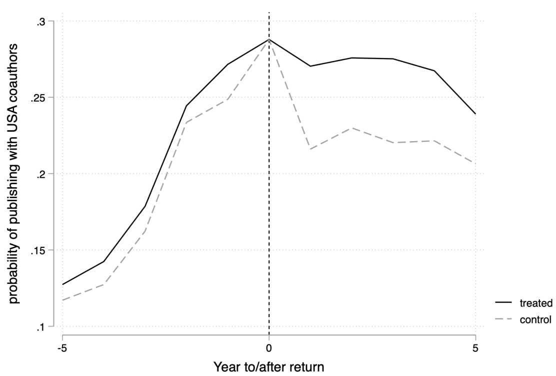
Figure A1: Publication Trends for Treated and Control Scientists



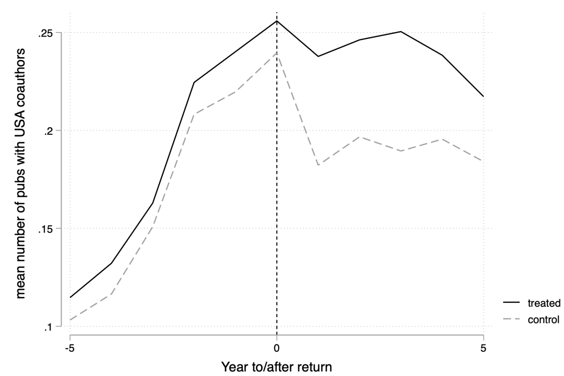
(a) Mean number of publications



(b) Any HIV publication



(c) Any publication with American-based coauthors



(d) Any new American-based coauthors

Appendix B: Additional Results

Table B1: Sensitivity Checks - Returnee Pre-Return Characteristics

Dept Var: num pubs	(1)	(2)	(3)	(4)	(5)
AFTER RETURN × RETURN INSTITUTION	0.062**	0.054	0.054*	0.0040	-0.013
	(0.034)	(0.033)	(0.032)	(0.0045)	(0.037)
AFTER RETURN × RETURN INSTITUTION × returnee productivity during fellowship		0.007			-0.0009
		(0.0081)			(0.0084)
AFTER RETURN × RETURN INSTITUTION × returnee PhD			0.13		0.19*
			(0.082)		(0.10)
AFTER RETURN × RETURN INSTITUTION × returnee from more advanced country				0.075	0.084
				(0.063)	(0.058)
Number of scientists	3,314	3,314	3,314	3,314	3,314
Number of scientists × year observations	36,454	36,454	36,454	36,454	36,454
Number of institutions	440	440	440	440	440

[a] Estimates stem from fixed effects ordinary least square specifications with dependent variables being inverse hyperbolic sine of counts of publications per scientist per year. All models incorporate a full suite of calendar year, career age and scientist fixed effects, and terms of the interaction of post-event and covariates are included.
[b] Heteroskedastic robust standard errors, clustered at the institution, are given in parentheses. *, **, *** denote statistical significance at p-values of 0.1, 0.05 and 0.01 respectively.

Chapter 4

4 Knowledge Production in Sub-Saharan Africa

Abstract

This paper enhances our understanding of the ideas production function for developing economies. It does so by considering international knowledge spillovers and cross-country teams as core determinants of technological catch-up. Using data of sub-Saharan African countries' scientific output between 1976 and 2016, I provide evidence for three main findings. First, the level of production of scientific output increases with the stock of ideas already discovered in a given country, as well as the level of human capital devoted to the scientific sector. Second, the level of production of scientific output is declining in the worldwide stock of ideas. That said, the level of production of scientific output of African countries increases with both the stock of ideas discovered in the ex-colonial power, and the levels of R&D funding of the ex-colonial power. This relationship is growing stronger over time, and is moderated by the size of the African country and their distance to the frontier. Third, the rate of collaboration between African and international scientists, particularly those from ex-colonial countries, is increasing over time. However, once this calendar trend is accounted for, international collaborations are more common for countries further behind the frontier. In an attempt to reconcile the findings, I find that the positive relationship between frontier country knowledge stocks and African publication output is moderated by the proportion of the African country scientific workforce that is involved in teamwork with the frontier country. I argue that international teamwork facilitates benefits from international knowledge spillovers and subsequent technological catch-up, particularly at earlier stages of development and for smaller countries. Overall, these findings are consistent with the concept that the rate of developing economy technological catch-up is associated with the production of knowledge in those developed countries with which they have relationships. Moreover, the findings suggest that knowledge, even that captured in scientific publications, is not easily accessible beyond these bilateral relationships, which has implications for programs and policies aiming to facilitate technological catch-up.

4.1 Introduction

Innovation is broadly regarded as a central driver of economic growth and the competitiveness of nations (Schumpeter 1942; Solow 1957; Abramovitz 1986; Romer 1990; Jones 1995). However, much of the world's new knowledge is produced in a handful of countries at the global technological frontier. In more recent years, the share of other countries in global knowledge production, in particular China and Korea, has grown significantly. But some of the world's countries, particularly those in sub-Saharan Africa, remain lagging far behind. International convergence in productivity, innovation output and economic growth ultimately depends on whether the knowledge generated in frontier countries can 'spill-over' to countries lagging behind.

Knowledge is inherently non-rival in its use, and Romer's model of endogenous innovation based growth is predicated on the notion that 'anyone engaged in research has free access to the entire stock of knowledge'. However, access to new knowledge is highly imperfect (Griliches 1957; Jaffe et al 1993; Audretsch and Feldman 1996; Jaffe and Trajtenberg 1999), and is likely to be particularly difficult for countries lagging behind the frontier (Cohen and Levinthal 1990; Jones 2014). Moreover, the contribution of international knowledge spillovers into production, i.e. foreign knowledge as a direct input into another country's production function, often embodied in capital goods, might differ from spillovers into research, i.e. foreign knowledge affecting another country's production by influencing its research productivity (Griliches 1979; Park 1995). Given challenges in diffusing knowledge around the globe, the latter is of critical importance to ensure long-run economic growth of a nation. This paper explores the role of international knowledge spillovers in innovative activities in countries lagging behind the frontier.

The extent to which countries lagging behind can utilize foreign knowledge to catch-up is a subject of much debate amongst researchers. On the one hand, the 'advantages of backwardness' (Gerschenkron 1962) imply that countries further behind can use knowledge generated elsewhere to skip some of the stages of development and catch-up. But on the other hand, scholars have argued a need for capabilities at the domestic level in order to benefit from catch-up growth (Abramovitz 1986). One approach to understanding the possibility of catch-up growth in scientific and innovative capacity is to focus on the domestic institutions required to support innovative activities, or national innovation systems (Freeman 1987; Lundvall 1992; Nelson 1993; Furman and Hayes 2004; Hu and Mathews 2005). While this literature has greatly advanced our understanding of the role of domestic institutions and capabilities in catch-up, studies in this tradition consider the country lagging behind as the unit of analysis, and implicitly assume that global knowledge production remains constant, or that it is independent of catch-up.

An alternative approach to the question is to document the implications of foreign knowledge stocks on domestic productivity. However, previous literature that has examined the role of international knowledge spillovers in innovation in countries lagging behind the frontier render mixed findings. Micro-level studies tracing knowledge flows into countries lagging behind find that knowledge, as measured through patent citations, can flow through social networks or migration of individuals (Singh 2005; Saxenian 2006; Oettl and Agrawal 2008; Kerr 2008; Choudhury 2015; Kahn and MacGarvie 2016). But connecting these knowledge flows to actual improvements in macro-level innovative performance has been more difficult. Several studies find limited evidence of a positive impact of trade or intellectual property protection, argued channels of knowledge spillovers, on developing country innovation outcomes (Park and Ginarte 1997; Schneider 2005; Kyle and McGahan 2012).

Two major challenges in these studies can go some way to explaining the mixed findings. First, studies of this kind typically use patents, or R&D spending, as a measure of innovation. The use of these measures could conflate innovation with competitive dynamics, trade relationships or pricing externalities (see Pavitt 1985 for an early summary of bias inherent in the use of patents as innovation indicators). Second, the latter studies tend to consider all foreign knowledge stocks as homogeneous, and assume that spillover benefits are determined by receiving country features. Given the micro-level evidence that knowledge flows through specific channels, it is possible that these findings could be masking heterogeneity in bilateral relationships and specific sending country knowledge stocks.

Fry (2020a) explores the impact of international knowledge spillovers on countries lagging behind the frontier using an innovation measurement that is arguably more independent from market forces and capital goods, scientific publications. The study examines the impact of the return home of U.S. trained scientists to African institutions on non-migrants affiliated with the institution they return to, and finds that non-migrant scientific output increases following the return event. The study finds empirical support for the hypothesis that knowledge spillovers from the specific U.S. training institution of the returning scientist are accessed through international teams. Whether this relationship holds at the aggregate level, and generalizes beyond returning scientists is the motivation behind this paper.

This paper estimates the shape of the scientific production function of countries lagging behind the frontier by differentiating between domestic and a variety of foreign knowledge stocks. I base the analysis on the conceptual framework outlined in Porter and Stern (2000), which evaluates the drivers of ideas sector productivity by allowing for separate contributions from the country-specific and rest-of-world knowledge stocks. This enables the evaluation of the relative importance played by international and domestic knowledge spillovers in fostering the production of new ideas in countries lagging behind the frontier.

To evaluate this empirically, I employ a panel dataset of 45 sub-Saharan African countries between 1976 and 2016. I explore the relationship between the flow of scientific publications for the African countries in the dataset in each year and the stock of both domestic and international knowledge. I document that (1) scientific productivity is increasing with domestic stocks of knowledge and human capital devoted to the scientific sector, but (2) declining in the worldwide stock of knowledge. To further explore the relationship between African publication output and foreign knowledge stocks I separate knowledge stocks by groups of countries. Foreign countries are grouped into three main categories as a preliminary exploration of mechanisms by which countries lagging behind the frontier benefit from international knowledge spillovers. First, in a test of whether geographic proximity influences benefits from international spillovers I measure the relationship between knowledge stocks in other African countries and African publication output. Second, in a test of whether a few frontier countries drive benefits from international knowledge spillovers I measure the relationship between OECD country knowledge stocks and African publication output. Third, in a test of whether bilateral relationships drive benefits from international knowledge spillovers and with enduring relationships between African and ex-colonial countries, particularly in science (Nagtegaal and de Bruin 1994; Staniland 1987), I measure the relationship between African publication output and the knowledge stocks in their ex-colonial power. I find that (3) scientific productivity is declining with respect to other African country knowledge stocks, and OECD country knowledge stocks, but (4) increasing with respect to ex-colonial country knowledge stocks and R&D spending, and that this relationship becomes more positive over the study time period.

The second objective of the study is to investigate the variation in the association between different foreign knowledge stocks and African publication output. Specifically, I hypothesize that international teamwork with specific frontier countries play a role in driving benefits from spillovers from those countries. With a rapid growth in international teams in global science, particularly those involving scientists from developing countries, teams can contribute to transferring knowledge, providing incentives to develop specialized skills, and ultimately enabling catch-up. I find descriptive evidence supporting this hypothesis. Once knowledge stocks in OECD countries are weighted by the proportion of scientific teamwork with that country, the relationship between the foreign knowledge stock and African scientific productivity becomes more positive. In exploring which African countries benefit from international teamwork, I find that frontier country-African country teams are more common for African countries with lower knowledge stocks, and that are smaller.

These findings are consistent with work that emphasizes the important role of ex-colonial countries in African science (Nagtegaal and de Bruin 1994; Staniland 1987). Although there is some debate about the nature and value of these relationships to African countries, I provide aggregate level evidence that African and ex-colonial countries knowledge production is at the least inter-related. More generally, relationships between frontier and African countries shape the extent to which African countries benefit from frontier country knowledge production in their own innovative activities.

The approach in this paper combines a formal model with insights from micro-level diffusion literature and accounts for some trends in relative innovation capabilities of countries lagging behind the frontier. It complements a vast body of literature that explores the determinants of national innovative output and technological catch-up from a single country perspective through highlighting the role of bilateral relationships and foreign innovative activities in domestic innovative capacity. Specifically the framework informs the finding that the innovative capacity of countries behind the frontier is increasing in the extent to which it has connections with high as opposed to low innovative frontier countries. This suggests that a focus on domestic institutions provides an incomplete explanation of the growth in innovative output of emerging economies in the last fifty years.

The remainder of the paper is structured as follows: Section 2 discusses prior literature on foreign knowledge and technological catch-up, Section 3 introduces the conceptual framework that drives the analysis. Section 4 outlines the empirical approach, data and measures. Section 5 presents results and Section 6 concludes and outlines implications of the findings.

4.2 Foreign Knowledge and Technological Catch-Up

A core economic issue is whether poor and rich countries will ultimately converge in the levels of income per capita and productivity. Since Solow (1957) and Abramovitz (1956) identified the importance of technology in economic growth, debates about the role of innovation in catch-up and convergence have intensified.

On the one hand, original models of technology driven economic growth categorized technology as a public good due to its non-rival nature. The conceptualization of technology, or knowledge, as freely available lends itself to potential 'advantages of backwardness' (Gerschenkron 1962) that implies that poorer countries can use knowledge generated elsewhere to skip some of the stages of development and catch-up.

But on the other hand, Gerschenkron (1962) himself noted that despite the promise of catch-up there are various challenges that could limit these theoretical benefits. Subsequent scholarship identified a need for capabilities at the domestic level in order to benefit from catch-up growth (Abramovitz 1986), and a tradition of research was born whereby scholars argue that creation and use of innovation is tied to economic institutions (Nelson and Winter 1982; Ames and Rosenberg 1963; see Fagerberg 1994 for a deeper review).

One progeny of this research is the national innovation systems approach, which focuses on relationships between firms, networks and institutions and their role in determining variation in innovative outcomes across countries (Freeman 1987; Lundvall 1992; Nelson 1993). While original research on the national systems of innovation had a bias towards more developed countries, and was traditionally more qualitative in methods, a small body of more recent quantitative evidence documents that domestic institutions are valuable in determining the rate of catch-up in innovative capacity in some of the rapidly growing Asian economies (Furman and Hayes 2004; Hu and Mathews 2005).

This nuanced perspective on the factors driving innovation and economic growth has developed alongside formal modeling of economic growth. In particular, models in the 'new growth theory' tradition attribute differences in economic development across countries to endogenous accumulation of knowledge (Romer 1990; Grossman and Helpman 1991). These models predict that a country's knowledge stock is of paramount importance in subsequent knowledge generation and that the level of the knowledge stock can go some way in explaining variation across countries in their ability to generate and use new innovations. While these models imply the critical role of factors affecting the mobility of knowledge across borders – which would lead to greater stocks – there is a level of abstraction that does not capture diffusion parameters.

These complementary perspectives that analyze single country models of economic growth and technological catch-up are supplemented by a deep and varied body of academic work documenting the implications of foreign knowledge stocks on domestic productivity, or international knowledge spillovers. These studies tend to fit into two main categories: micro studies that trace knowledge flows directly through measuring citations between countries and studies that take a macro perspective evaluating the role of knowledge flows, as proxied through trade, FDI or disembodied spillovers, on productivity at the country, or country-industry level.

Zvi Griliches' (1957) pioneering work demonstrating the diffusion of hybrid corn seeds stimulated a variety of studies focused on understanding the parameters affecting the diffusion of knowledge. Research tracing flows of knowledge using citation patterns document that flows of knowledge are mediated in large part by geographic distance (Jaffe et al 1993;

Zucker et al 1998; Jaffe and Trajtenberg 1999). This suggests that for countries lagging behind that are also far away from frontier countries, access to frontier knowledge can be extremely challenging. However, recent research suggests that knowledge can flow into countries and regions lagging behind through other channels, including social networks, migration and ethnic diaspora in frontier countries (Singh 2005; Saxenian 2006; Oettl and Agrawal 2008; Kerr 2008; Choudhury 2015). While the evidence that knowledge can flow into countries lagging behind the frontier through specific channels has been at the micro-level of the piece of knowledge, studies connecting knowledge flows to actual productivity gains have largely been at the level of the country, or country-industry.

Econometric research on international knowledge spillovers at the macro level was spurred by Coe and Helpman (1995), who find large spillovers from foreign knowledge stocks on domestic total factor productivity (TFP) in 22 developed countries, where foreign knowledge stocks are constructed using the weighted sum of trade partners' cumulative R&D spending. Extending this, Coe et al (1997) test for the presence of North-South R&D spillovers using a sample of 77 developing countries, again finding substantial spillovers. In a complementary approach, Eaton and Kortum (1996; 1999) model R&D and the diffusion of knowledge using international patenting rates, and document significantly larger productivity growth from foreign R&D than domestic R&D for OECD countries. Subsequent research explores the role of trade, FDI and communication patterns in international knowledge spillovers (see Keller 2004 for a comprehensive summary of this empirical literature). Although there is some level of disagreement on the precise measures, the literature converges on the idea that knowledge produced elsewhere can have an impact on productivity in a focal country, and that access to foreign knowledge can be shaped by bilateral relationships.

Beyond the general relationship between international knowledge stocks, bilateral relationships and productivity, there is a role for domestic capabilities in determining the level of benefit from international knowledge spillovers. In theory, the higher the level of capabilities, or absorptive capacity, of a country the more it will benefit from foreign R&D (Cohen and Levinthal 1989), and empirical work confirms this idea. Blomstrom et al (1994) explore the role of FDI on productivity growth and find that the impact of FDI on growth is greater the higher the level of development of the host country. Similarly, Coe et al (2009) find that domestic institutions mediate the extent of benefits from knowledge spillovers from trade, and Eaton and Kortum (1996) find that a country's level of education plays a role in the ability to absorb foreign ideas. Kerr (2008) explores the transfer of knowledge through ethnic communities in the U.S. to their home countries and finds that manufacturing productivity improvements arising from knowledge transfer through ethnic networks are lower for less developed countries. To the extent that domestic capabilities shape benefits from international knowledge spillovers, catch-up is not automatic and could present particular challenges for countries lagging farthest behind the frontier.

Furthermore, spillovers into research may be different from spillovers into production. The distinction between the two is important. First, if benefits from knowledge spillovers are embodied in capital, such as a piece of equipment, the spillover may take the form of a pricing externality, would have different consequences for endogenous growth (Griliches 1979). Second, the direction of the spillover may be different. While benefits from foreign knowledge spillovers can provide a valuable input into the production of new knowledge and technological catch-up (Romer 1990), the creation of foreign knowledge can raise the bar for new knowledge production making it harder for countries to catch-up (Jones 2009; Bloom et al 2017). The relative strength of these opposing effects will determine the possibility of convergence in innovation output.

Empirical work exploring this tension renders mixed findings. While Park (1995) finds a positive correlation between foreign R&D spending – weighted by bilateral technological distance – and domestic R&D spending in OECD countries, Kyle et al (2017) report a negative relationship between foreign R&D spending and domestic R&D spend in neglected disease R&D. Similarly, Porter and Stern (2000) examine innovation output as measured through patents of OECD countries as a function of domestic and foreign knowledge stocks, and find a negative relationship between global knowledge stocks and domestic innovative output, concluding that the crowding out effect of foreign knowledge stocks dominates any benefits from knowledge spillovers.

While these studies use samples of OECD, or frontier, countries to explore international knowledge spillovers into research, it is possible that countries lagging behind the frontier benefit even less from knowledge spillovers. Innovation at the global frontier requires a level of skill and codified, or tacit, knowledge that might not be available in developing countries. Empirical evidence supports this theory. Schneider (2005) examines the relationship between knowledge spillovers, proxied by trade flows, and innovation outcomes in both developed and developing countries, finding a much weaker relationship between the two in developing countries. Similarly, while studies exploring the consequences of trade liberalization policies find a positive impact in developed countries, the results are more bleak in developing countries, documenting a limited impact of trade policies on developing country knowledge production (Park and Ginarte 1997; Kyle and McGahan 2012).

To advance our understanding of the determinants of technological catch-up in countries lagging behind the frontier I take a different approach from the extant literature. In doing so I extend and bring together the two lines of research on international knowledge spillovers discussed above that have mostly developed in isolation. Through evaluating the production function of ideas focusing specifically on the role of foreign knowledge this paper is designed to contribute to our understanding of the role of international knowledge spillovers in the production of new knowledge in countries lagging behind the frontier.

4.3 Conceptual Approach

The approach to assessing national innovative productivity is based on the ideas production function articulated by Romer (1990) and Jones (1995) and extended by Porter and Stern (2000) to include international knowledge spillovers. I first describe a production function for brand new ideas for countries lagging behind the frontier:

$$\dot{A}_{jt} = \delta H_{jt}^{\gamma} A_{jt}^{\theta} S_{-jt}^{\sigma} \tag{1}$$

where for each country j in year t, $\dot{A}_{jt}$ represents the flow of brand new ideas, and H_{jt} is the quantity of human capital devoted to the ideas-producing sector. The flow of new knowledge is described as a function of the stock of previous knowledge generated in the country, A_{jt}, and in the rest of the world, S_{-jt}, which is global knowledge stocks across countries other than (j). Conceptually, together knowledge stocks (A + S) have opposing effects: they could facilitate the production of new knowledge (standing on the shoulders of giants), or they may make the production of new ideas more difficult (fishing-out effect). The sign on θ and σ nets out the opposing effects. If they are positive, the standing on the shoulders dominates. If they are negative, the fishing-out dominates.

The magnitude of both θ and σ are crucial to the debate about convergence in ideas production. In the original Romer (1990) model of ideas production, $\theta = \sigma = 1$. This implies that a percentage increase in the stock of ideas anywhere in the world result in a proportional increase in the productivity of the ideas sector. Under this assumption, policies that shift the scientific workforce (even temporarily) would permanently shift the growth rate of ideas production. Jones (1995) relaxed the assumption of proportional increase and suggested that the strength of spillovers could be less than proportional. This implies that with weaker spillovers in ideas if there are no increases to the scientific workforce, growth rate in ideas production will be zero. Similar to Porter and Stern (2000) I further extend the Romer-Jones models to allow for both concavity of ideas production as well as heterogeneity in the strength of spillovers from domestic and foreign knowledge stocks.

The extent to which countries lagging behind the frontier benefit from international knowledge spillovers and subsequently catch-up depends on the level of accessibility of foreign knowledge stocks. The ideas production function can be enhanced by incorporating parameters for the accessibility of specific foreign knowledge stocks. Superior access and use of foreign knowledge would imply greater values of σ. Most earlier research attempting to understand international knowledge flows focuses on access to foreign knowledge and subsequent spillovers through channels such as bilateral trade, foreign direct investment (FDI), technological proximity, migration and social networks. The emphasis of this paper is on how the stock of international ideas can be accessed through international teamwork.

International Teamwork Driving Benefits from Knowledge Spillovers

The rate of international collaborative teams in science and innovation, particularly those involving innovators from countries lagging behind the frontier, has grown significantly over the last fifty years (see Wagner and Leydesdorff 2005 for a summary).

For innovators in countries lagging behind, collaborative teamwork with frontier countries can help in two main ways. First, knowledge and resources flow through ties (Granovetter 1973; Singh 2005; Agrawal et al 2008; Fry 2020b), particularly collaborative ties (Singh 2005). Second, knowledge workers have endogenously acquired skills, and international teamwork can provide complementary skills to stimulate domestic workers into specializing and participating in frontier innovation (Wuchty et al 2008; Jones 2014). It follows, therefore, that an increase in cross-country teamwork with a frontier country implies an increase in the benefits from knowledge spillovers from that country. Similarly, an increase in knowledge stocks of a frontier country raises the benefits to countries that are engaged in cross-country teamwork with them.

Given the intuition that an increase in cross-country teamwork between a frontier country and a country lagging behind increases the potential benefits from knowledge spillovers, I propose a functional form based on the Romer-Jones-Porter/Stern model allowing for heterogeneous benefits from foreign knowledge stocks as a function of bilateral teamwork. I draw from classical models estimating the benefits of international knowledge spillovers in production started with Coe and Helpman (1995) in which foreign knowledge stocks enter separately and are weighted by the level of access to the stock. The weight in these macro-studies is traditionally a measure of bilateral trade or foreign direct investment. I extend the production function of ideas through weighting each foreign country's stock of knowledge with a measure for the extent to which the African country is engaged in collaborative teamwork with that country. Consider the national ideas production function for country j,

$$\dot{A}_{jt} = \delta H_{jt}^{\gamma} A_{jt}^{\theta} Sw_{-jt}^{\sigma'} \tag{2}$$

where again $\dot{A}_{jt}$ represents the flow of brand new ideas, and H_{jt} is the quantity of human capital devoted to the ideas-producing sector, and the flow of new knowledge is described as a function of the stock of previous knowledge generated in the country, A_{jt}. Unlike equation (1), Sw_{-jt} is defined as the sum of foreign countries' knowledge stocks, with each country's stock i at time t weighted by the percentage of publications coauthored between country i and j at time t (similar to the weighting scheme in Coe and Helpman 1995 used for import shares). In notational terms this is $\sum_{i \neq j}(s_{it} \times$ (fraction of coauthored publications between i and j in year t)) whereby s_{it} is country i's cumulative number of publications at time t. If international teamwork is associated with benefits from international knowledge spillovers I would expect $\sigma' > \sigma$.

There is also an argument to be made that the effects of teams with frontier country scientists may differ over time, and across countries lagging behind. From the perspective of the country lagging behind, teams with frontier countries form, and enable improvements to innovative output if the benefit outweighs the cost. Costs can considered as predominantly those associated with communication. With declining communication costs over time, it is likely that international collaborative teams, and thus international knowledge spillover benefits, are increasing over time (Ding et al 2010). Benefits to international teamwork are likely to be associated with both how far a country is from the frontier (Branstetter et al 2018), as well as how large the scientific workforce is within a country (Jones 2014). That being said, I would expect the following relationship to hold: *International teams spanning frontier and countries lagging behind are most valuable for countries in later time periods, and for countries that are smaller and lagging further behind the global frontier.* Thus the expectation is that σ' is greater in later time periods, for smaller countries and for countries lagging further behind the frontier.

4.4 Empirical Approach

I derive the link between the national ideas production function and international knowledge. In doing so, I develop an econometric model to estimate the parameters of (1) and (2). There are several issues that the econometric model must account for. I take each issue in turn to arrive at an estimating equation that can be applied to data on innovation in countries lagging behind the frontier.

First, the model must account for the relationship between innovation measures and market dynamics. It is standard to measure innovation output at the country level using counts of international (usually U.S.) patents. However, it is extremely difficult to distinguish true innovation from market dynamics using these measures. Given that patenting often captures innovations embodied in capital goods, the use of patents as a measure for both innovation output and knowledge stocks can be misleading. Country X could be applying for U.S. patents more frequently than country Y because it would like to trade with the U.S., independent of whether it is more innovative than country Y. Moreover, for countries lagging behind the frontier, where resources are scarce, the cost associated with patenting in the U.S. can be prohibitive. It is also very difficult to disentangle knowledge spillovers from pricing externalities when using patents as a measure of international knowledge stocks. Therefore I use scientific publications as a measure of innovation in this study for two main reasons: (1) scientific publications are less likely to be contaminated by market dynamics than patents are; (2) knowledge captured in scientific publications is less likely to be 'wrapped up' in a capital good than knowledge captured in patents.

Second, there is likely to be variation across countries, and across time in terms of propensity to publish. Thus an identification challenge in this study arises if the same observed or unobserved factors affect either a country's publication output across years, or both the focal and the foreign countries' publication output in a given year. As a source of statistical identification, I employ a panel dataset and take advantage of both cross-sectional and time series variation in estimating the parameters of the equations above. I employ both country and year fixed effects in the majority of specifications which accounts for any underlying heterogeneity at the country level, as well as any trends that could affect the whole world simultaneously. This allows the interpretation of the coefficients as within country changes accounting for overall global trends.

Third, it is important to capture the lag in time necessary from starting to innovate to observation of a scientific publication. I impose a lag of two years between observed publications and the variables associated with national ideas production.

Assuming that the terms are complementary with one another, denoting the natural logarithm of A as LA, I suggest the specification to estimate the parameters in equation (1):

$$L\dot{A}_{jt+2} = \delta_t + \lambda_j + \gamma LH_{jt} + \theta LA_{jt} + \sigma LS_{-jt} + e_{jt} \tag{3}$$

And the specification in equation (2) reduces to the following form:

$$L\dot{A}_{jt+2} = \delta_t + \lambda_j + \gamma LH_{jt} + \theta LA_{jt} + \sigma' LSw_{-jt} + e_{jt} \tag{4}$$

Using a least squares method, the log-log form of these specifications allow many of the variables to be interpreted in terms of elasticities. I employ a panel dataset over a time period of 41 years for 45 African countries to estimate the parameters in the equations (3) and (4). The remainder of the section describes the data used and how the measures are constructed.

Data and Measures

The empirical setting for this study is African science. Although countries lagging behind the frontier are growing at a rate relatively faster than those at the frontier (OECD countries) in the period 1976-2016, compared to countries with similar initial levels, and particularly per capita, both the productivity and growth of countries in sub-Saharan Africa is extremely low (Table 1; Figure 1). Furthermore, domestic funding for science and R&D in African countries is extremely low, and policies and programs supporting science and R&D are mostly driven by foreign partners. This provides a motivation to examine the extent to which African countries are able to benefit from international knowledge spillovers.

The data consists of a novel dataset of publication activity from 1976 through 2016 for all sub-Saharan African countries, excluding South Africa.[17] Table 2 provides definitions and sources, and Table 3 presents summary statistics. To estimate the production function

[17] South Africa is at a very different stage of development in terms of its innovative capacity from the remainder of the continent, and so is excluded from the analysis.

for new ideas, the data include (a) the flow of publications in each African country in each year, (b) measures of the factor inputs into ideas production, (c) publication level measurements on team composition.

Publications. The principal dependent variable, $Publications_{jt+2}$, is the number of publications (articles) with authors affiliated with a given country j in year (t+2) as found in the Elsevier Scopus database. The average number of publications produced by an African country in a given year between 1976 and 2016 in the sample is 177, with a standard deviation of 523. Distinguishing those publications in regional versus international journals (a rough proxy for local versus global frontier work) gives an average of 158 international journal publications a year per country. The bottom five rows of Table 1 illustrate that this variation across African countries in growth rates is not driven by population differences alone, and not only does the absolute level of publication output differ across the continent, but also the growth rate. The empirical work explores this heterogeneity amongst African countries in terms of structural factors.

Factor Inputs for the Ideas Production Function. I estimate the sensitivity of the production of ideas to human capital and both domestic and foreign knowledge stocks. I measure human capital using the number of unique names associated with a focal country affiliation captured in publications in year t. There is a lack of data on innovation in general, particularly labor market data, in sub-Saharan Africa and so this gives a consistent measure across countries. The key assumption I am making is that each active scientist is publishing at least one publication a year. This assumption is hard to verify, but this can be considered the threshold of the definition of a scientist in this study. This averages at 228 scientists in a country year in the sample (89% of which publish in international journals).

The domestic stock of knowledge is defined as each country's cumulative publication stock, which is the sum of all publications for all years prior (since 1976[18]). I extend the ideas production function to separate out the effects of the within-country publication stock from international publication stocks. The world publication stock is the cumulative sum of publications over all countries in a given year, but from the point of view of the

[18]Chosen for reliability of the data as there are few publications in Elsevier Scopus database earlier than 1976.

country j, the relevant concept is 'Rest of World' publication stock (world publication stock - own publication stock). I also separate foreign knowledge stocks into categories of foreign countries to distinguish between different mechanisms that could drive international knowledge spillovers. Three main groups of countries are used: (a) the knowledge stock for the rest of Africa (excluding South Africa), which is calculated as the total number of publications with African affiliated authors minus the focal countries publications; (b) OECD country knowledge stocks; and (c) the knowledge stock of the focal country's (most recent) ex-colonial power.

A second measure of knowledge stocks is given by R&D funding levels. Data on R&D funding by African countries is limited, however in reality expenditures are negligible and so can be ignored. Data on R&D funding at the OECD country level is measured using 2010 USD for consistency and gathered from the OECD stat. database. I incorporate the sum of funding across all OECD countries in a given year, and by groups similar to that outlined above for publication stocks in the framework.

Descriptive statistics on knowledge stocks and R&D funding across all ex-colonial OECD countries are provided in Table 5. Two things to note: (1) the scientific output in these OECD countries is orders of magnitude greater than that in African countries; and (2), there is significant variation in scientific output and R&D funding amongst the ex-colonial OECD countries. This variation is exploited in several of the specifications.

Weights for International Knowledge Stocks. The latter part of the analysis incorporates measurements for the composition of teams of African country publications. I capture teamwork with OECD scientists through extracting affiliation countries of coauthors in the focal African country's publications. Table 4 provides summary statistics across all publications across the African countries in the time period. 42% of all African publications are co-authored with any OECD country based scientists, and 14% with ex-colonial based scientists, with the rate of collaborations increasing over time (Figure 2). Again, there is considerable variation in collaboration rates with OECD countries across the African countries. Over the sample period, the rate of collaboration ranges from as low as 0% to 100% in some country years. In the earliest time period (1986-1995), almost half of the countries had a rate of less between 25-35%, while less than a quarter of countries had more than 75%. To calculate the weights used in the model, for a given focal country year, I calculate the proportion of publications coauthored with scientists with affiliations from each OECD country in turn. OECD country knowledge stocks in that year are then multiplied by that proportion, and this weighted knowledge stock is summed over all OECD countries.

4.5 Results

The results proceed in several steps. First, I provide evidence for the association between foreign knowledge and domestic knowledge production in countries behind the frontier. In particular I find that there is heterogeneity in the sensitivity of domestic knowledge production to foreign knowledge stocks according to the 'sending' country. Specifically, publication output in African countries is positively associated with ex-colonial country knowledge stocks, but negatively associated with the rest of the global knowledge stock. Second, I find descriptive support for the hypothesis that international teams are a way to access foreign knowledge stocks, and to benefit from international knowledge spillovers.

The Ideas Production Function

The econometric analysis applies the specification in equation (3) to the core dataset of 1,845 observations. Table 6 presents several models providing the primary production function results, reproducing the Romer-Jones-Porter/Stern ideas production function model. Namely, a log-log regression of the flow of publications in an African country on the amount of labor in the scientific sector, a measure of the stock of knowledge, and some controls for country and year. The baseline specification (6-1) includes fixed effects for both country and year. The results show that number of scientists, and knowledge stock in a country have a significant and economically important relationship with publication output. The coefficient on LH_{jt}, or domestic human capital, implies that a 10% increase in scientists is associated with a 2.6% increase in publication output, and the coefficient θ on LA_{jt}, or country level knowledge stock, implies that a 10% increase in the knowledge stock is associated with a 3.5% increase in publication output. Importantly, future production is concave in both past stock and scientific workforce, contrary to Romer (1990) that presumed that the production of new ideas is growing proportionally with the stock of ideas discovered in the past. This suggests that for the sample of African countries in this study ideas are getting harder to find, and that this effect dominates an effect of 'standing on the shoulders' of prior knowledge, limiting the possibility of endogenous growth in ideas production. This result is consistent with the documented rates of growth in publication output in Table 1: that countries with lower levels of absolute production of ideas per capita experience higher rates of growth in production of ideas.

In model (6-2) I relax the controls accounting for heterogeneity across years and countries by including just a baseline year dummy, a year trend and a baseline country proxy (log GDP in 1976). While there is a decline in the sensitivity of knowledge production to the scientific workforce, there is a substantial increase in the sensitivity to the knowledge stock. The increase in θ depends mostly on the exclusion of country-specific effects. Thus the support in favor of a 'standing on the shoulders' mechanism driving knowledge production is stronger when relying on cross-country variation. Model (6-3) confirms this interpretation. The inclusion of country fixed effects, a baseline year dummy, and a year trend into the model again reduces the sensitivity to the knowledge stock. But the difference in θ in models (6-1) and (6-3) also suggests that variation over time that isn't captured in a year time trend can explain a significant portion of the sensitivity to knowledge stocks. Together, these results suggest that publication productivity in African countries is mostly driven by time-invariant differences across countries, but there is also a reasonable amount of variation that can be explained by changes over time. Another point worth noting, is that even with only a handful of regressors, nearly all of the variance in publication rates ($R^2 > 0.9$) is explained.

Model (6-2) also shows a downward evolution in publication productivity over time (although this evolution is small at around 0.6% a year). This result, together with the positive and significant coefficient on country knowledge stock, suggests that time-series variation in publication productivity is a tension between the positive impact of a country's knowledge stock, and an overall negative effect year to year. So while convergence amongst countries behind the frontier seems possible, whether the negative trend of productivity is an African specific trend or a global trend will determine the rate of convergence between countries at the frontier and those lagging behind. The remainder of the table explores this through highlighting the role of international knowledge spillovers.

In (6-4, 6-5, and 6-6) I follow the specification suggested in equation (3) and include variables for the rest of the world knowledge stock. Note that the coefficient on S_{-jt} is not separately identified from individual time effects, because own knowledge stock plus rest of the world stock is constant across countries in a given year. Therefore, because the level of publications across countries grow at a relatively constant rate across years, I capture variance through the inclusion of an overall time trend as opposed to year fixed effects. The results show that publication output of African countries is negatively correlated with knowledge stocks in the rest of the world (ideas are getting harder to find) (6-4), and the rest of Africa (albeit less of a negative relationship) (6-5), and OECD countries knowledge stock (6-6). This suggests that the impact of international knowledge production raising the bar for new knowledge production dominates a positive impact of international knowledge spillovers.

Separating OECD knowledge stocks into distinct categories and including them in the preferred specification with country and year fixed effects (which allows for more flexibility in the data), (6-7) illustrates that African publication output is negatively correlated with OECD country (excluding their ex-colonial power) knowledge stocks. However,(6-8) reports that a 10% increase in the knowledge stock of the ex-colonial power amounts to a 2.4% increase in the output of the African country. This is almost the same elasticity to its own country knowledge stock. In contrast to previous literature that finds no or negative effects of foreign knowledge on domestic productivity, this captures the evidence of positive spillovers from specific countries.

Table 7 explores the relationship between foreign R&D spending and African publication output. (7-1) provides the baseline specification including year and country fixed effects. (7-2) includes covariates for the sum of OECD country (excluding the specific ex-colonial country) R&D spending in a given year to the main specification, and reports that the relationship between OECD spending and African publication output is negative. Column (4) includes R&D spending of the ex-colonial country into the specification, and reports that the relationship to publication output in Africa is positive. The coefficient on $LRnD_{-jt}$ implies that a 10% increase in ex-colonial country R&D funding is associated with a statistically significant 3.9% increase in African publication output. In order to understand whether the observed relationship between international funding and African publication output is driven by indirect changes to the international knowledge stocks, models (7-3) and (7-5) control for publication stock in OECD countries and the ex-colonial country, respectively. With the coefficient on the R&D funding in ex-colonial country decreasing significantly with the addition of this control variable, I interpret that the majority of the association between African publication output and ex-colonial R&D funding is via the ex-colonial country knowledge stock.

Considering that countries may go through different stages of development and that stage of development as well as temporal trends may influence the association between foreign knowledge stocks and African publication output, Table 8 reports results for separate time periods. A few trends merit discussion: (1) the association between African publication flows and domestic knowledge stocks increases over time, (2) the association between OECD country (excluding ex-colonial power) knowledge stocks and African publication output become more negative over time, and (3) the association between knowledge stocks in ex-colonial powers and African publication output becomes more positive over time. This suggests that while there is in theory an overall raising of the bar to produce new knowledge, the benefits from knowledge spillovers from both own-country knowledge stocks and ex-colonial power stocks increase in more recent time periods.

In order to better understand which kinds of countries benefit more from foreign knowledge stocks, I first split the sample into above and below median sized as per population, and run the main specification on each sample separately (Table 9). Consistent with Coe and Helpman (1995), the relationship between ex-colonial knowledge stocks and scientific productivity is driven by smaller African countries, while larger countries rely more on their own stock than foreign stocks. Importantly, the relationship between OECD country (excluding ex-colonial power) knowledge stocks and African publication output is more negative for larger countries. This points to the idea that the tension between benefiting from international knowledge stocks and ideas getting harder to find is moderated by country level features, a finding of which is explored in the latter part of the paper. Second, I split the sample into those further behind the frontier, and closer to the frontier according to their publication flow per capita in a given year. Table 10 reports the results of the specification on these separate samples, and highlight that the association between foreign knowledge stocks and publication output is stronger for those countries lagging further behind. This is surprising, given that prior literature documents the importance of absorptive capacity or country level capabilities to benefit from international knowledge spillovers. However, if we expect that there are specific channels by which countries behind the frontier can benefit from international knowledge spillovers, and countries lagging behind invest relatively more in those channels, this could go some way to reconciling this seemingly contradictory evidence.

The remainder of this paper explores whether there is an association between international teamwork, foreign knowledge stocks and African publication output.

The Role of International Teams in Catching-Up

First I explore whether the relationship between ex-colonial countries and African publication output is related to a greater propensity to collaborate with the ex-colonial country than other OECD countries. Figure 3 shows that for African countries, collaborative teams amongst their ex-colonial country are much more prevalent than teams amongst other countries. Table 11 provides the regression counterpart to this figure. I use a dataset at the level of African/OECD country pair year, and regress the probability that the African country coauthors with that specific OECD country on a dummy for whether the OECD country is the ex-colonial power. Column (1) illustrates that African countries have a significant propensity to collaborate with an OECD country given that it is the ex-colonial power. I run a variety of specifications with different control variables to verify that this relationship isn't driven by features of the ex-colonial country as opposed to a true propensity to collaborate with the specific ex-colonial country. In column (2) I account for the size of the OECD country in the model, to verify that the observed relationship is not driven by the possibility that ex-colonial countries are larger than other OECD countries. The results remain robust. Column (3) adds in controls for the size of the knowledge stock of the OECD countries to account for the possibility that ex-colonial powers are leaders in terms of scientific output and that this could be driving the observed relationship. The results remain robust. Finally in column (4) I incorporate year fixed effects into the specification to verify that the relationship isn't driven by macro-level trends. Again the results are robust. Thus I can say with some confidence that African countries have a higher propensity to collaborate with ex-colonial country scientists than any other OECD country scientists.

Table 12 uses a dataset at the African publication level and estimates the propensity for collaboration with OECD scientists as a function of time, country size and country knowledge stock. The table shows that OECD coauthors are more likely to be found on an African publication in later time periods, and on publications emanating from smaller African countries, and countries lagging further behind the frontier (as shown by the negative and statistically significant coefficient on log knowledge stock, even when controlling for population). The same trends are found for ex-colonial country coauthors, although the relationship between coauthoring patterns and both time and country features are less strong. As African countries have a long history of relationships with ex-colonial countries in a variety of formats, the reduced reliance of the formation of collaborative relationships on global trends and country features is not surprising. That being said, these results are consistent with the theory that international teams, particularly between lagging and frontier countries are both a way to access frontier knowledge, as well as to complement domestic knowledge workers.

Building on these observed trends of coauthoring patterns, I seek to explore if there is a relationship between international knowledge spillovers and international teamwork. In doing so I calculate bilateral team shares for all African-OECD country pairs and incorporate them as weights of the OECD country knowledge stocks in the main regression model. Table 13 reports the results from the specification in equation (4). Column (1) and (2) provide the baseline results of the association of African publication output and OECD country knowledge stocks. As described earlier, African publication output is negatively correlated with OECD country (excluding ex-colonial power) knowledge stocks. A 10% increase in OECD country knowledge stocks correlates with a 2.8% decline in African publication output. Once the OECD country knowledge stocks are weighted for the proportion of publications that include collaborators from the specific OECD country (column 3), this relationship becomes less negative. I interpret this as the positive benefits from international spillovers increasing, while the negative effects of 'fishing out' of global knowledge stocks remain constant. Once similarly weighted ex-colonial country knowledge stocks are included (column 4) the relationship becomes even less negative – a precisely estimated zero. In this instance the positive benefits of international spillovers counterbalance the negative effects of fishing out. Columns (5) and (6) explore a model excluding year fixed effects, and instead include a year time trend and a dummy for the baseline year. This allows me to separately identify the relationship between African publication output and unweighted knowledge stocks from all OECD countries (column 5). Similar to the baseline specification, there is a negative relationship between African publication output and OECD country knowledge stocks. However, once OECD country knowledge stocks are weighted as before in column (6), this negative relationship again becomes a precisely estimated zero. Again - I interpret this as international teamwork facilitating the access of international knowledge stocks, and once this access is accounted for, the negative effect from fishing out is counterbalanced by the positive impact of accessing foreign knowledge through international collaborative teams.

These results of the elasticity of domestic innovation to foreign knowledge stocks are surprisingly small given that Coe et al (1997) who use a similar weighting scheme find that total factor productivity in developing countries is positively and significantly related to R&D in their industrial country trade partners, and to their imports of equipment from industrial countries. This implies two possible interpretations of the relative role of foreign knowledge in developing country innovation as compared to production: (1) the fishing out effect in innovation is significant, or (2) for domestic production of ideas to benefit from foreign knowledge production an additional level of tacit knowledge is required beyond that which is required to use equipment or machinery in production. Distinguishing between the two is beyond the scope of this paper, but is an important avenue for future research.

I explore potential mechanisms driving this observed association between foreign knowledge stocks and African publication output. In Table 14 I estimate the baseline production function with a variety of different outcomes. First, I assess whether the association between ex-colonial knowledge stocks and domestic productivity is direct or indirect. Direct would imply that the African country and the ex-colonial country co-produce knowledge together at a greater rate with greater knowledge stocks, while indirect implies that there is a knowledge spillover to non co-produced output. Table 14 column (2) reports the association between ex-colonial knowledge stocks and the production of papers coauthored with OECD based scientists, and column (3) publications coauthored with ex-colonial scientists. The negative relationship in both specifications implies that co-production is not driving the results. This suggests an indirect knowledge spillover is more salient. Second, I assess whether the African country is benefiting through reproducing the same knowledge, or producing brand new to the world knowledge. Using publications in regional journals as a proxy for reproduction (brand new to the region, but not necessarily brand new to the world), column (4) illustrates that again, this is not driving the result, suggesting that the African countries are producing more brand new to the world knowledge.

In sum, I find support for the theoretical expectation that international teams, particularly those spanning frontier countries and countries lagging behind the frontier, are a way to access international knowledge and complementary skills required to achieve growth in knowledge production. Moreover, I find that international teams spanning frontier and developing countries are most valuable for countries lagging further behind the global frontier, and for smaller countries.

Robustness

In this paper I document a previously unexplored phenomenon: the positive association between particular foreign knowledge stocks, and technological catch-up in ideas production, and I find supportive evidence that one factor driving this relationship is the participation in international teams. Although I attribute this phenomenon to a mechanism of knowledge transfer and co-specialization benefits, there are other possible explanations. I explore a few of these below and in Table 15.

There could be the concern that the relationship between ex-colonial and African country productivity is driven by heterogeneity in underlying institutions in some African countries (Acemoglu et al 2001), which is correlated with the colonialists who tend to produce and transport more knowledge abroad. In order to consider if this is driving the results, I first use country fixed effects to account for underlying institutions. Second I use a proxy for underlying institutions that should be unrelated to the knowledge producing behavior of the ex-colonial power: white settlement (Acemoglu et al 2001). A dummy variable at the level of the African country is generated that takes the value of 1 if there was significant white settlement in colonial times, which in turn led to stronger institution building by the settlers. Interacting the dummy for white settlement with the knowledge stocks from the ex-colonial power I am able to deduce whether countries with better institutions are driving the result. Table 15 column (2) finds that those countries without white settlement actually drive the results – the opposite of what we would expect if ex-colonial power institutional development is driving the results.

Another possibility is that the results are being driven by other bilateral relationships that happen to be correlated with team composition. This could include trade relationships. I add a control variable of ex-colonial knowledge stocks weighted by the proportion of imports in the focal African country year from the ex-colonial country (similar to Coe and Helpman 1995; Coe et al 1997) in column 3. The association between African publication output and ex-colonial knowledge stocks remains robust after controlling for imports.

Finally, I consider the possibility that results are driven by common language. The concern is that I could be picking up an association between ex-colonial country knowledge stocks and African productivity due to their shared language as opposed to bilateral relationships. To verify that this isn't the case I measure the association between African country publication output and knowledge stocks in OECD countries that share a common language, excluding the ex-colonial power, in Table 15 column 4. The negative coefficient on the knowledge stock in these OECD countries suggests that it is not common language driving the results.

4.6 Conclusion

This paper contributes to a long line of research investigating factors affecting national innovative output, and technological catch-up of countries lagging behind. While previous literature has focused on country level investments (Furman and Hayes 2004; Hu and Matthews 2004), I explore the role of international knowledge spillovers. Using an approach combining formal economic modeling and insights from more micro-level traditions, I examine the production of scientific output in 45 sub-Saharan African countries during the years 1976-2016, and the relationship between publication output at the African country level and domestic and foreign knowledge stocks.

A number of observations emerge: (1) scientific productivity is increasing with domestic stocks of knowledge and human capital devoted to the scientific sector, but (2) declining in the worldwide stock of knowledge. To explore the relationship between foreign knowledge stocks and African publication output I separate foreign knowledge stocks by groups of countries: (a) knowledge stocks in other African countries; (b) knowledge stocks in OECD countries; and (c) knowledge stocks in ex-colonial powers. Each of these categories highlights a slightly different hypothesized mechanism by which countries can benefit from international knowledge spillovers, namely (a) geographic proximity, (b) frontier country dominance, and (c) bilateral relationships. I find that (3) scientific productivity is declining with respect to other African country knowledge stocks, and OECD country knowledge stocks, but (4) increasing with respect to their ex-colonial country knowledge stocks and R&D spending, and that this relationship is growing over the study time period.

In the latter part of the paper I hypothesize that international teams play a role in driving benefits from international knowledge spillovers, and that this could go some way to explaining the observations above. In an exploration of this hypothesis I weight OECD country knowledge stocks by the proportion of bilateral teamwork between the focal African country and the OECD country. Inclusion of the sum weighted stocks of foreign knowledge results in the relationship between the weighted knowledge stock and African scientific output becoming more positive. Moreover these relationships are more positive for African countries that engage in relatively more international teamwork, namely those that are smaller and further behind the frontier. Although preliminary, these results are consistent with the theory that countries lagging behind the frontier can use international teamwork as a way to transfer knowledge and build complementary skills.

One of the more surprising findings is the negative association between African publication output and foreign knowledge stocks in OECD countries other than their ex-colonial power. This finding provides support for complementary interpretations: that foreign knowledge production raises the bar for new ideas production, and that international knowledge spillovers are relatively weak. Scientific progress therefore is a tension between these two competing dynamics. The positive association between African publication output and their ex-colonial country knowledge stocks illustrates a tip in this balance. Given that the positive association does not hold for countries other than the ex-colonial power, the only explanation is that relationships with ex-colonial powers strengthen benefits from international knowledge spillovers. This is consistent with the voluminous body of research that highlights the difficulties in global knowledge diffusion (Jaffe et al 1993; Zucker et al 1998; Jaffe and Trajtenberg 1999). At the same time, it can help to reconcile the inconclusive evidence on the relevance of foreign knowledge for domestic innovation (Park and Ginarte 1997; Porter and Stern 2000; Kyle and McGahan 2012; Fry 2020). In particular, the results highlight that the capacity to produce ideas depends critically on the existence of channels by which foreign knowledge is accessed.

These findings inform a large debate on the feasibility of convergence. The 'advantages of backwardness' (Gerschenkron 1962), and the promise of technological catch-up, particularly in an increasingly globalized world, have not been realized in practice. Today there remain just under 750 million people living on less than $2 a day, more than half of whom live in sub-Saharan Africa.[19] The critical feature of any model of convergence is that ideas generated in the world are freely available. However, scholars have made significant progress on documenting the limits to knowledge availability. Therefore the feasibility of convergence, and the conditions under which it could happen, is an open question. This paper contributes to this debate in a few ways. First, I examine the role of international knowledge in knowledge generation in countries lagging behind the frontier, which is distinct from the role of international knowledge in production, and arguably could lead to greater levels of convergence if pricing externalities exist in the latter (Griliches 1979). Second, I measure innovation using scientific publications in order to reduce the bias associated with strategic and trading relationships between countries inherent in the use of patents as innovation output. And third, I consider bilateral relationships, specifically the prevalence of cross country teams, as a way to access foreign knowledge and benefit from international knowledge spillovers. The findings from this study are able to reconcile some of the mixed evidence between micro- and macro- level studies in this area, and highlight that foreign knowledge stocks do matter for technological catch-up, but that there are

[19] https://www.worldbank.org/en/topic/poverty/overview

limitations in the extent to which countries lagging behind the frontier are able to access and use foreign knowledge.

More broadly, the approach helps to explain some of the variation in innovative capacity amongst the world's poorest countries. Specifically, innovative capacity of a country lagging behind the frontier is increasing in the extent to which it has connections with high- as opposed to low- innovative frontier countries. I show that domestic institutions provide at best an incomplete explanation of the growth in innovative output of emerging economies in the last fifty years. This implies that programs and policies supporting the development of innovative capacity through strengthening cross-country relationships could have a large positive impact.

This study has a few major limitations. The main limitation of the study is the descriptive nature of the data. While the use of country and year fixed effects attempt to account for some unobserved bias that could be driving the results, the findings should be interpreted as descriptive. Future research should seek to unpack some of these findings using experimental methods to tease out causal processes driving the observed relationship between foreign knowledge stocks and the scientific performance of countries lagging behind. Other major limitations of the study relate to the measurement of innovation and inputs into the ideas production function. I use publication records as a proxy for performance. Whether this is a true reflection of innovative performance is unclear. Similarly, the use of publication data to gather statistics on the scientific workforce in Africa is problematic, particularly if this does not capture scientists who are less productive. Given data constraints these seemed like the best measures available that would be consistent across countries, but again, future research can consider alternative metrics to measure innovative performance.

This paper raises more questions than it answers: Is there variation by resource intensity of the field? Are similar relationships observed amongst multi-national firms? Do other drivers of foreign knowledge access, such as the diaspora or migration, explain variation in innovative output? Future work should seek to explore the role of international knowledge to further our understanding of the development of innovative capacity in countries lagging behind the frontier.

References

Abramovitz, M., 1956 *Resource and Output Trends in the United States Since 1870* American Economic Review 46(2): 5-23

Abramovitz, M., 1986 *Catching Up, Forging Ahead and Falling Behind* Journal of Economic History 46: 385-406

Agrawal, A., Kapur, D., McHale, J., 2008 *How Do Spatial and Social Proximity Influence Knowledge Flows? Evidence From Patent Data* Journal of Urban Economics 64: 258-269

Ames, E., Rosenberg, N., 1963 *Changing Technological Leadership and Industrial Growth* The Economic Journal 73(289): 13-31

Audretsch, D.B., Feldman, M.P., 1996 *R&D Spillovers and the Geography of Innovation and Production* The American Economic Review 86(3): 630-640

Bloom, N., et al., 2017 *Are Ideas Getting Harder to Find?* No. w23782. National Bureau of Economic Research

Blomstrom, M., Lipsey, R.E., Zejan, M., 1994 *What Explains the Growth of Developing countries?* Convergence of Productivity: Cross-National Studies and Historical Evidence: 243-259

Branstetter, L., Glennon, B., Jensen, J.B., 2018 *Knowledge Transfer Abroad: The Role of U.S. Inventors Within Global R&D Networks* Working Paper

Choudhury, P., 2015 *Return Migration and Geography of Innovation in MNEs: A Natural Experiment of Knowledge Production by Local Workers Reporting to Return Migrants* Journal of Economic Geography 16: 585-610

Cohen, W.M., Levinthal, D.A., 1990 *Absorptive Capacity: A New Perspective on Learning and Innovation* Administrative Science Quarterly: 128-152

Coe, D.T., Helpman, E., 1995 *International R&D Ppillovers* European Economic Review 39: 859-87

Coe, D.T., Helpman, E., Hoffmaister, A.W. 1997 *North-South R&D Spillovers* The Economic Journal 107(440): 134-149

Coe, D.T., Helpman, E., Hoffmaister, A.W., 2009 *International R&D Spillovers and Institutions* European Economic Review 53(7): 723-741

Ding, W.W., et al., 2010 *The Impact of Information Technology on Academic Scientists' Productivity and Collaboration Patterns* Management Science 56(9): 1439-1461

Eaton, J., Kortum, S., 1996 *Trade in Ideas: Patenting and Productivity in the OECD* Journal of International Economics 40(3): 2-8

Eaton, J., Kortum, S., 1999 *International Technology Diffusion: Theory and Measurement* International Economic Review 40: 537-570

Fagerberg, J.,1994 *Technology and International Differences in Growth Rates* Journal of economic Literature 32(3): 1147-1175

Freeman, C., 1987 *Technology Policy and Economic Performance: Lessons from Japan* Pinter, London, UK

Fry, C.V., 2020a *Building Bridges: The Impact of Return Migration of African Scientists* Working paper

Fry, C.V., 2020b *Viral Privilege: Evidence From the Ebola Epidemic* Working paper

Furman, J.L., Hayes, R., 2004 *Catching up or Standing Still? National innovative productivity among 'follower' countries, 1978-1999* Research Policy 33: 1329-1354

Gerschenkron, A., 1962 *Economic Backwardness in Historical Perspective* Belknap Press of Harvard University Press, Cambridge, MA

Granovetter, M.S., 1973 *The Strength of Weak Ties* American Journal of Sociology LXXIII, 1360-1380

Griliches, Z., 1957 *Hybrid Corn: An Exploration in the Economics of Technological Change* Econometrica 25(4): 501-522

Griliches, Z., 1979 *Issues in Assessing the Contribution of R&D to Productivity Growth* Bell Journal of Economics 10: 92-116

Grossman, G., Helpman, E., 1991 *Innovation and Growth in the Global Economy* MIT Press, Cambridge, MA

Hu, M-C, Mathews, J.A., 2005 *National Innovative Capacity in East Asia* Research Policy 34(9): 1322-1349

Jaffe, A.B., Trajtenberg, M., Henderson, R., 1993 *Geographic Localization of Knowledge Flows as Evidenced by Patent citations* Quarterly Journal of Economics CVIII: 577-598

Jaffe, A.B., Trajtenberg, M., 1999 *International Knowledge Flows: Evidence from Patent Citations* Economics of Innovation and New Technology 8.1-2: 105-136

Jones, C., 1995 *R&D Based Models of Economic Growth* Journal of Political Economy 103: 739-784

Jones, B., 2009 *The Burden of Knowledge and the Death of the Renaissance Man: Is Innovation Getting Harder?* The Review of Economic Studies 76(1)

Jones, B.F., 2014 *The Knowledge Trap: Human Capital and Development Reconsidered* Working Paper

Kahn, S., MacGarvie, M., 2016 *Do Return Requirements Increase International Knowledge Diffusion? Evidence from the Fulbright Program* Research Policy 45(6): 1304-1322

Keller, W., 2004 *International Technology Diffusion* Journal of Economic Literature 42(3): 752-782

Kerr, W.R., 2008 *Ethnic Scientific Communities and International Technology Diffusion* The Review of Economics and Statistics 90(3): 518-537

Kyle, M.K., McGahan, A.M., 2012 *Investments in Pharmaceuticals Before and After TRIPS* Review of Economics and Statistics 94(4): 1157-1172

Kyle, M.K., Ridley, D.B., Zhang, S., 2017 *Strategic Interaction Among Governments in the Provision of a Global Public Good* Journal of Public Economics 156: 185-199

Lundvall, B.A., 1992 *National Systems of Innovation: Towards a Theory of Innovation and Interactive Learning* London, UK, Pinter

Nagtegaal, L.W., De Bruin, R.E., 1994 *The French Connection and Other Neo-Colonial Patterns in the Global Network of Science* Research Evaluation 4(2) : 119–127

Nelson, R.R., Winter, S., 1982 *An Evolutionary Theory of Economic Change* Harvard University Press, Cambridge, MA

Nelson, R.R. (Ed.), 1993 *National Innovation Systems: A Comparative Analysis* Oxford University Press, New York, NY

Oettl, A., Agrawal, A., 2008 *International Labor Mobility and Knowledge Flow Externalities* Journal of International Business Studies 39(8): 1242-1260

Park, W.G., 1995 *International R&D Spillovers and OECD Economic Growth* Economic Inquiry 33:571-91

Park, W.G., Ginarte, J.C., 1997 *Intellectual Property Rights and Economic Growth* Contemporary Economic Policy 15(3): 51-61

Pavitt, K., 1985 *Patent Statistics as Indicators of Innovative Activities: Possibilities and Problems* Scientometrics 7.1-2: 77-99

Porter, M.E, Stern, S., 2000 *Measuring the 'ideas' Production Function: Evidence from International Patent Output* NBER Working Paper 7891

Romer, P., 1990 *Endogenous Technological Change* Journal of Political Economy 98: S71-S102

Saxenian, A., 2006 *The New Argonauts: Regional Advantage in a Global Economy* Cambridge, MA: Harvard University Press

Schneider, P.H., 2005 *International Trade, Economic Growth and Intellectual Property Rights: A Panel Data Study of Developed and Developing Countries* Journal of Development Economics 78(2): 529-547

Schumpeter, J., 1942 *Capitalism, Socialism and Democracy* 2nd ed.George Allen & Unwin, Ltd., London

Singh, J., 2005 *Collaborative Networks as Determinants of Knowledge Diffusion Patterns* Management Science 51(5): 756-770

Solow, R.M., 1957 *Technical Change and the Aggregate Production Function* The Review of Economics and Statistics: 312-320

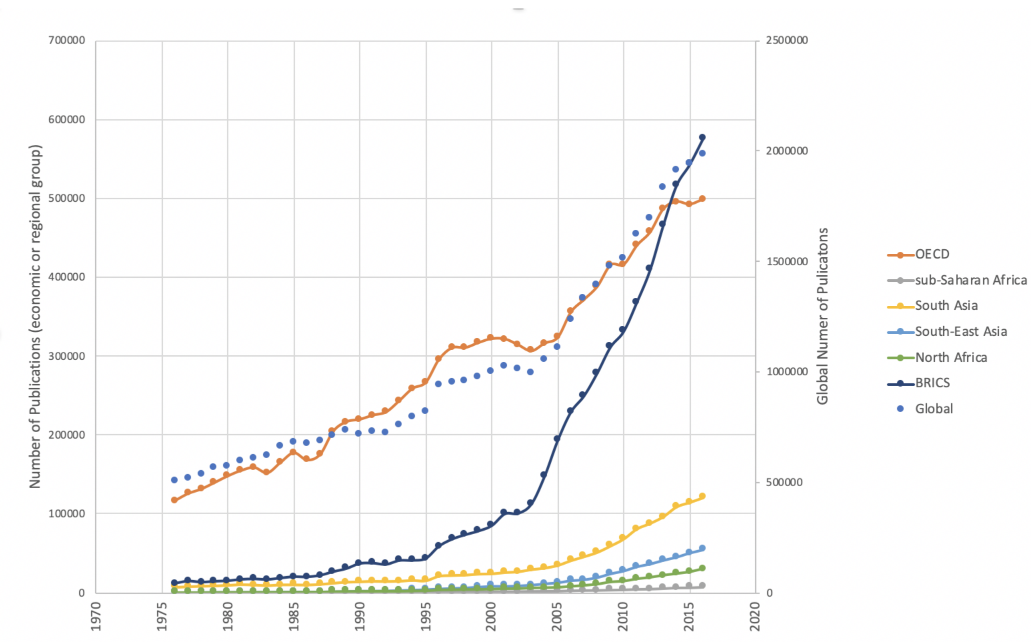
Staniland, M. 1987) *Francophone Africa: The Enduring French Connection* Annals of the American Academy of Political and Social Sciences 489 : 51-62

Wagner, C.S., Leydesdorff, L., 2005 *Network Structure, Self-organization, and the Growth of International Collaboration in Science* Research Policy 34(10): 1608-1618

Wuchty, S., Jones, B.F., Uzzi, B., 2007 *The Increasing Dominance of Teams in Production of Knowledge* Science 316.5827: 1036-1039

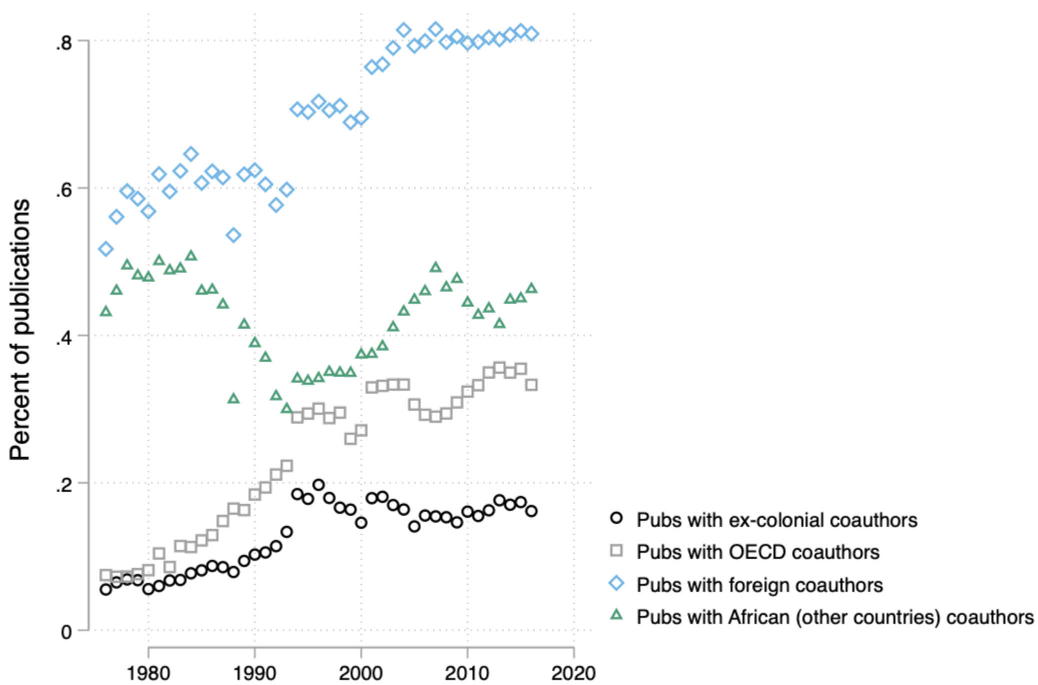
Zucker, L., Darby, M., Brewer, M., 1998 *Intellectual Capital and the Birth of U.S. Biotechnology Enterprises* American Economic Review 88(1): 290-306

Figure 1: Publication Trends Worldwide



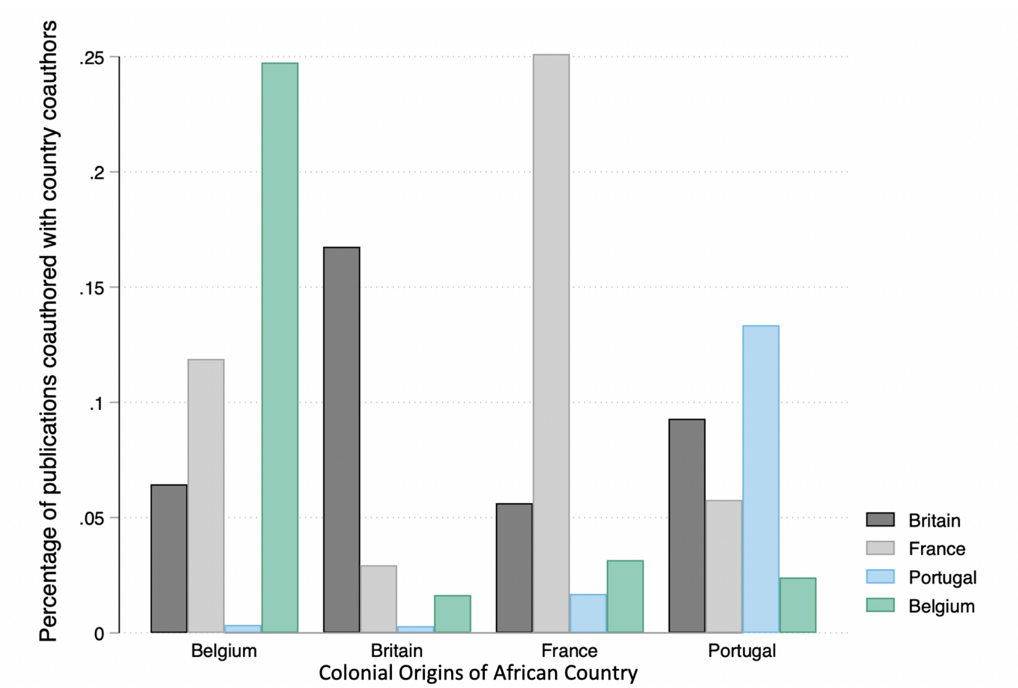
Note: The absolute number of publications per year is plotted for economic and geographic regions of the world.

Figure 2: Average Percentage of Publications with Foreign Teams Across Sub-Saharan Africa



Note: The average number of publications across all African countries co-authored with each foreign group in a given year is plotted.

Figure 3: Percentage of Publications with ex-Colonial Country Coauthors



Note: The graph shows the average percentage of publications per year co-authored with specific OECD country scientists, aggregated by colonial origin of African country. On the x-axis is the colonial origin, over which African countries are grouped - on the y-axis is the average percentage of articles coauthored with given country over African countries in the group between 1976 and 2016. For example, the first dark gray bar provides the proportion of publications with British coauthors (1976-2016) amongst African countries that were under Belgium rule.

Table 1: Country statistics: averages for 10- and 20-year periods (scientific publications)

	Publications per year			Publications per year per capita (millions)			Annual growth rate (%)		
	1976-1996	1996-2006	2006-2016	1976-1996	1996-2006	2006-2016	1976-1996	1996-2006	2006-2016
OECD	188,864	317,192	437,510	238.5	349.6	464.0	4.9	2.7	4.1
USA	185,461	274,740	375,902	769.5	967.4	1208.7	4.4	2.1	3.8
France	26,701	49,710	71,499	468.5	808.7	1094.6	5.2	4.0	4.6
UK	46,305	76,458	108,446	814.1	1290.2	1715.4	4.2	2.7	4.8
BRICS	26,340	113,285	389,167	13.0	44.2	136.3	8.8	16.8	10.5
China	7,463	69,275	268,577	7.0	54.6	199.3	41.2	22.0	10.2
South Africa	2,679	4,684	10,772	80.7	102.9	205.2	4.7	4.9	9.8
South Asia	11,969	27,622	79,548	11.6	19.5	47.6	5.8	9.6	12.0
India	673,476	1,028,613	1,658,067	847.1	957.7	1,320.1	3.3	3.9	6.2
Sri Lanka	133	290	740	8.2	15.3	36.3	3.8	8.8	10.9
South-East Asia	2,137	9,432	33,261	5.3	17.9	55.3	9.9	11.9	13.8
Thailand	482	2,223	7,500	9.3	35.3	111.9	10.2	15.8	11.1
Vietnam	76	427	1,992	1.2	5.3	22.4	20.1	14.2	18.5
North Africa	1,866	5,813	17,919	32.9	76.7	205.3	7.8	9.5	13.7
Algeria	119	593	2,722	5.2	19.1	73.6	14	17.5	18.4
Sub-Saharan Africa (excl South Africa)	783	1,926	5,001	1.91	3.1	6.0	6.9	9.0	11.4
Kenya	365	727	1,927	17.2	21.8	43.6	5.7	5.3	11.0
Malawi	41	140	443	4.5	12.0	29.1	10.1	11	15.8
Zambia	103	122	329	13.5	11.4	22.8	9.8	3.4	11.4
Mali	25	79	199	3.1	6.8	12.7	23.9	11.3	7.6
Guinea	8	18	58	1.4	2.1	5.4	24.1	42.9	14.5

Note: Statistics aggregated by groups of countries in terms of economic classification or regional (not necessarily mutually exclusive), as well as by selected individual countries are provided.

Table 2: Variables and Definitions

Variable	Full Variable Name	Definition	Source
Innovative Output			
A_{jt+2}	Publications in year t + 2	Number of published articles with an author affiliated with country j in year t+2	Elsevier scopus database
Human Capital			
H_{Ajt}	Number of scientists in year t	Number of unique names affiliated with country j in published articles in year t	Elsevier scopus database
Knowledge Stocks			
A_{jt}	Stock of publications up until year t	Cumulative number of published articles (from 1976) with an author affiliated with country j up until year t	Elsevier scopus database
S_{-jt}	Stock of rest of the world's publications up until year t	Cumulative number of global published articles (from 1976), minus any publications with an author affiliated with country j, up until year t. Additional variables created for the cumulative numbers of OECD country publications minus ex-colonial power publications, African publications and ex-colonial power publications	Elsevier scopus database
Sw_{-jt}	Weighted stock of rest of the world's publications up until year t	Sum of cumulative publications of each country i other than j up until time t, weighted by proportion of coauthored publications between i and j in time t	Elsevier scopus database
Foreign R&D Spending			
RnD_{-jt}	R&D spending in OECD countries in year t	Total gross domestic expenditure across OECD countries on R&D in year t (1996-2016[20])	OECD Stat database
Controls			
POP_{jt}	Population in year t	Population of country j in year t	World Bank
GDP_{jt}	GDP in year t	GDP of country j in year t	World Bank

Table 3: Summary Statistics (African country year level)

	mean	median	std. dev.	min.	max.
Scientific Output					
Number of publications	177.5	33	523.6	0	8466
Number of international journal publications	158.2	31	459.4	0	7648
Number of regional journal publications	19.3	2	70.7	0	891
Number of neglected tropical disease publications	37.74	7	88.36	0	1057
Number of publications coauthored with OECD scientists	79.14	15	186.18	0	2663
Number of publications coauthored with ex-colonial scientists	26.47	6	59.24	0	922
Number of publications coauthored with OECD scientists by country	3.31	0.43	8.71	0	112
Number of publications coauthored with other African country scientists	75.86	3	460.7	0	8466
Number of publications coauthored no non-African scientists	88.97	12	356.9	0	5600
Scientific Workforce					
Number of publishing authors	228.8	36	693.4	0	9457
Number of publishing authors in international journals	204.1	33	608.0	0	8623
Number of publishing authors only in regional journals	24.7	1	93.9	0	1224
Knowledge Stock					
Stock of publications	1761.2	286	5616.3	0	83363

Note: This table provides details on the 45 African countries over 41 years (1976-2016) in the sample.

Table 4: Summary Statistics (African publication level)

	mean	median	std. dev.	min.	max.
Publication Quality					
Number of citations	14.03	5	47.5	0	7057
Journal impact factor	0.80	0.74	0.94	0	14.7
Team Composition					
Team size	5.14	4	16.39	1	3581
Dummy international team	0.74	1	0.44	0	1
Dummy other African country coauthors	0.38	0	0.49	0	1
Dummy just other African country coauthors	0.27	0	0.44	0	1
Dummy OECD coauthors	0.42	0	0.49	0	1
Dummy ex-colonial coauthors	0.14	0	0.35	0	1

Note: This table provides details on the full sample of 329,720 publications with authors originating from any of the sample African countries over the time period 1976-2016.

Table 5: Summary Statistics (OECD ex-colonial country year level)

Country (ex-colonial to # of African countries)	mean	median	std. dev.	min.	max.
OECD Knowledge Flows (1976-2016)	283,348	295,306	114,500	116,439	498,218
UK (16)	69,463	71,575	28,565	31,215	130,281
France (15)	43,715	45,586			
Portugal (5)	4,396	2,351	4,975	178	16,634
Belgium (4)	10,216	9,552	6,100	3,760	23,383
Germany (2)	62,310	63,636	27,911	28,651	119,435
Italy (2)	33,650	32,965	18,230	10,966	74,148
Spain (1)	24,538	21,653	20,528	2,344	67,489
USA (1)	256,046	262,458	89,303	120,581	428,763
OECD R&D Funding (in millions USD) (2007-2015)					
UK (16)	35	35	2	33	38
France (15)	46	46	1.6	43	48
Portugal (5)	3.3	3.2	0.32	2.9	3.8
Belgium (4)	9.0	8.9	1.1	7.2	10
Germany (2)	84	84	6.7	73	96
Italy (2)	22	22	9.1	21	23
Spain (1)	16	16	7.9	15	17
USA (1)	398	391	23	373	441

Note: This table provides details on the full OECD countries considered 'ex-colonial' in the sample. Descriptive statistics of knowledge flows are given as averages for the country over the full sample years 1976-2016. Funding descriptives are given similarly, but for the time period 1996-2016 as the data permits.

Table 6: Scientific Production Function in Sub-Saharan Africa

Dependent Variable = log(publication count)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
LH_{jt}	0.26***	0.15***	0.32***	0.28***	0.32***	0.28***	0.26***	0.29***
	(0.017)	(0.0094)	(0.018)	(0.017)	(0.018)	(0.017)	(0.017)	(0.018)
LA_{jt}	0.35***	0.63***	0.18***	0.32***	0.20***	0.33***	0.35***	0.30***
	(0.028)	(0.014)	(0.025)	(0.028)	(0.026)	(0.028)	(0.028)	(0.029)
LS_{-jt}				-0.47***				
				(0.047)				
LS_{-jt} rest of Africa					-0.11***			
					(0.031)			
LS_{-jt} OECD						-0.48***		
						(0.044)		
LS_{-jt} OECD countries (excl ex-colonial power)							-0.028**	
							(0.014)	
LS_{-jt} ex-colonial power								0.24***
								(0.047)
Controls								
Log population	0.37**	0.41**	0.33*	0.40**	0.34**	0.40**	0.38**	0.34**
	(0.12)	(0.13)	(0.13)	(0.13)	(0.13)	(0.013)	(0.12)	(0.12)
Year		-0.006***	0.027***	0.049***	0.036***	0.051***		
		(0.0016)	(0.0043)	(0.0047)	(0.0049)	(0.0047)		
Log GDP 1976		0.019***						
		(0.0036)						
Year fixed effects	yes						yes	yes
Country fixed effects	yes		yes	yes	yes	yes	yes	yes
R^2	0.95	0.90	0.94	0.94	0.94	0.94	0.95	0.95
Adjusted R^2	0.94	0.90	0.93	0.93	0.93	0.94	0.94	0.94
Number of observations	1,845	1,845	1,845	1,845	1,845	1,845	1,845	1,845

Note: Estimates stem from fixed effects ordinary least square specifications with dependent variables being log of counts of outcomes of country j in year t+2. Robust standard errors are given in parentheses. *, **, *** denote statistical significance at p-values of 0.1, 0.05 and 0.01 respectively.

Table 7: International Funding and Scientific Productivity in Sub-Saharan Africa

Dependent Variable = log(publication count)	(1)	(2)	(3)	(4)	(5)
LH_{jt}	0.18***	0.18***	0.18***	0.19***	0.20***
	(0.031)	(0.030)	(0.030)	(0.031)	(0.031)
LA_{jt}	0.12**	0.12**	0.12**	0.095*	0.062
	(0.057)	(0.057)	(0.058)	(0.057)	(0.059)
$LRnD_{-jt}$ OECD countries (excl ex-colonial power)		-6.63	-6.82*		
		(4.03)	(4.11)		
$LRnD_{-jt}$ ex-colonial power				0.39**	0.11
				(0.12)	(0.14)
LS_{-jt} OECD countries (excl ex-colonial power)			-0.034		
			(0.14)		
LS_{-jt} ex-colonial power					0.31**
					(0.14)
Controls					
Log population	0.34	0.31	0.32	0.29	0.27
	(0.23)	(0.23)	(0.24)	(0.23)	(0.23)
Year fixed effects	yes	yes	yes	yes	yes
Country fixed effects	yes	yes	yes	yes	yes
R^2	0.95	0.95	0.95	0.95	0.95
Adjusted R^2	0.94	0.94	0.94	0.94	0.94
Number of observations	990	990	990	990	990

Note: Estimates stem from fixed effects ordinary least square specifications with dependent variables being log of counts of outcomes of country j in year t+2. Robust standard errors are given in parentheses. *, **, *** denote statistical significance at p-values of 0.1, 0.05 and 0.01 respectively.

Table 8: International Spillovers and Scientific Productivity in Sub-Saharan Africa Changes Over Time

Dependent Variable = log(publication count)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
		1976-1995			1996-2005			2006-2016	
LH_{jt}	0.20***	0.20***	0.21***	0.068	0.065	0.080*	-0.021	-0.026	-0.0097
	(0.028)	(0.028)	(0.029)	(0.048)	(0.048)	(0.048)	(0.041)	(0.041)	(0.041)
LA_{jt}	0.11**	0.11**	0.10**	0.19*	0.22*	0.15	0.23	0.18	0.076
	(0.045)	(0.045)	(0.045)	(0.12)	(0.12)	(0.12)	(0.17)	(0.17)	(0.18)
LS_{-jt} OECD countries		-0.0058			0.34			-1.51	
(excl ex-colonial power)		(0.019)			(0.25)			(0.97)	
LS_{-jt} ex-colonial power			0.17			0.56**			0.63**
			(0.13)			(0.24)			(0.26)
Controls									
Log population	0.48*	0.48*	0.49**	0.34	0.23	0.28	0.31	0.35	0.18
	(0.025)	(0.25)	(0.25)	(0.57)	(0.58)	(0.57)	(0.59)	(0.59)	(0.59)
Year fixed effects	yes	yes	yes	yes	yes	yes	yes	yes	yes
Country fixed effects	yes	yes	yes	yes	yes	yes	yes	yes	yes
R^2	0.95	0.95	0.95	0.97	0.97	0.97	0.96	0.96	0.96
Adjusted R^2	0.94	0.94	0.94	0.96	0.96	0.96	0.95	0.95	0.95
Number of observations	900	900	900	450	450	450	495	495	495

Note: Estimates stem from fixed effects ordinary least square specifications with dependent variables being log of counts of outcomes of country j in year t+2. Robust standard errors are given in parentheses. *, **, *** denote statistical significance at p-values of 0.1, 0.05 and 0.01 respectively.

Table 9: International Spillovers and Scientific Productivity in Sub-Saharan Africa: by African country size

Dependent Variable = log(publication count)	(1)	(2)	(3)	(4)	(5)	(6)
	below median sized countries			above median sized countries		
LH_{jt}	0.18***	0.17***	0.25***	0.33***	0.34***	0.33***
	(0.025)	(0.025)	(0.031)	(0.026)	(0.027)	(0.027)
LA_{jt}	0.38***	0.38***	0.32***	0.43***	0.40***	0.42***
	(0.036)	(0.036)	(0.039)	(0.057)	(0.063)	(0.059)
LS_{-jt} OECD countries (excl ex-colonial power)		-0.027*			-3.52	
		(0.015)			(2.90)	
LS_{-jt} ex-colonial power			0.27***			0.071
			(0.074)			(0.066)
Controls						
Log population	-0.27*	-0.24	-0.26*	1.12***	1.08***	1.09***
	(0.16)	(0.16)	(0.16)	(0.24)	(0.024)	(0.24)
Year fixed effects	yes	yes	yes	yes	yes	yes
Country fixed effects	yes	yes	yes	yes	yes	yes
R^2	0.91	0.91	0.91	0.95	0.95	0.95
Adjusted R^2	0.90	0.90	0.90	0.94	0.94	0.94
Number of observations	901	901	901	944	944	944

Note: Estimates stem from fixed effects ordinary least square specifications with dependent variables being log of counts of outcomes of country j in year t+2. Robust standard errors are given in parentheses. *, **, *** denote statistical significance at p-values of 0.1, 0.05 and 0.01 respectively.

Table 10: International Spillovers and Scientific Productivity in Sub-Saharan Africa: by African country distance to frontier

Dependent Variable = log(publication count)	(1)	(2)	(3)	(4)	(5)	(6)
	countries lagging behind			countries forging ahead		
LH_{jt}	0.21***	0.21***	0.21***	0.29***	0.29***	0.31***
	(0.024)	(0.024)	(0.024)	(0.021)	(0.021)	(0.025)
LA_{jt}	0.27***	0.27***	0.25***	0.32***	0.32***	0.30***
	(0.037)	(0.037)	(0.038)	(0.035)	(0.035)	(0.039)
LS_{-jt} OECD countries (excl ex-colonial power)		0.025			0.0050	
		(0.046)			(0.016)	
LS_{-jt} ex-colonial power			0.12**			0.083
			(0.058)			(0.071)
Controls						
Log population	0.99***	0.99***	0.93***	0.053	0.053	0.061
	(0.020)	(0.020)	(0.20)	(0.11)	(0.11)	(0.11)
Year fixed effects	yes	yes	yes	yes	yes	yes
Country fixed effects	yes	yes	yes	yes	yes	yes
R^2	0.93	0.93	0.93	0.98	0.98	0.98
Adjusted R^2	0.92	0.92	0.92	0.97	0.97	0.97
Number of observations	943	943	943	902	902	902

Note: The sample is split into countries above median publication flows 'forging ahead' and below median 'lagging behind' in each year, and the main specification is estimated separately on these sub-samples. Estimates stem from fixed effects ordinary least square specifications with dependent variables being log of counts of outcomes of country j in year t+2. Robust standard errors are given in parentheses. *, **, *** denote statistical significance at p-values of 0.1, 0.05 and 0.01 respectively.

Table 11: Propensity for African Scientists to Collaborate with Ex-Colonial Country Scientists

Dependent Variable = coauthored publication proportion				
	(1)	(2)	(3)	(4)
Ex-colonial country	0.18***	0.18***	0.18***	0.17***
	(0.0017)	(0.0017)	(0.0017)	(0.0017)
L(OECD country population)		0.27***	0.31***	0.34***
		(0.0062)	(0.0064)	(0.019)
L(OECD country knowledge stock)			-0.002***	-0.002***
			(0.000)	(0.000)
Controls				
Year fixed effects				yes
R^2	0.19	0.22	0.22	0.25
Adjusted R^2	0.19	0.22	0.22	0.25
Number of observations	44,832	44,832	44,832	44,832

Note: Estimates stem from fixed effects ordinary least square specifications with dependent variables being log of counts of outcomes for a given country j/OECD country pair in year t+2 Robust standard errors are given in parentheses. *, **, *** denote statistical significance at p-values of 0.1, 0.05 and 0.01 respectively.

Table 12: Team Composition in Sub-Saharan Africa

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dependent variable	Publication with OECD coauthor				Publication with ex-colonial coauthor			
Year	0.0090***				0.0023***			
	(0.000)				(0.000)			
Log Population		-0.097***		-0.023***		-0.055***		-0.053***
		(0.000)		(0.0012)		(0.000)		(0.000)
Log Knowledge Stock			-0.036***	-0.095***			-0.019***	-0.0065***
			(0.000)	(0.0011)			(0.000)	(0.000)
Controls								
Year fixed effects				yes				yes
Number of observations	328,248	328,248	328,248	328,248	328,248	328,248	328,248	328,248

Note: Estimates stem from fixed effects ordinary least square specifications with dependent variables being dummy outcome per publication across the entire sample of African countries between 1976 and 2016. Robust standard errors are given in parentheses. *, **, *** denote statistical significance at p-values of 0.1, 0.05 and 0.01 respectively.

Table 13: International Spillovers and Scientific Productivity in Sub-Saharan Africa: with weighted foreign knowledge stocks

Dependent Variable = log(publication count)	(1)	(2)	(3)	(4)	(5)	(6)
LH_{jt}	0.26***	0.26***	0.26***	0.26***	0.30***	0.32***
	(0.017)	(0.017)	(0.017)	(0.017)	(0.018)	(0.018)
LA_{jt}	0.35***	0.35***	0.37***	0.35***	0.20***	0.19***
	(0.028)	(0.028)	(0.030)	(0.029)	(0.025)	(0.027)
LS_{-jt} OECD countries (excl ex-colonial power)		-0.028**				
		(0.014)				
LS_{-jt} OECD					-0.079***	
					(0.014)	
LSw_{-jt} OECD countries (excl ex-colonial power)			0.0052			
			(0.0034)			
LSw_{-jt} OECD countries				0.00084		-0.0033
				(0.0037)		(0.0038)
Controls						
Log population	0.37**	0.38***	0.34**	0.37**	0.36**	0.32**
	(0.12)	(0.12)	(0.12)	(0.12)	(0.13)	(0.13)
Year					0.032***	0.027***
					(0.0044)	(0.0043)
Year fixed effects	yes	yes	yes	yes		
Country fixed effects	yes	yes	yes	yes	yes	yes
R^2	0.95	0.94	0.95	0.95	0.94	0.94
Adjusted R^2	0.94	0.94	0.94	0.94	0.93	0.93
Number of observations	1,845	1,845	1,845	1,845	1,845	1,845

Note: Estimates stem from fixed effects ordinary least square specifications with dependent variables being log of counts of outcomes of country j in year t+2. Robust standard errors are given in parentheses. *, **, *** denote statistical significance at p-values of 0.1, 0.05 and 0.01 respectively.

Table 14: International Spillovers and Variations in Publication Outcomes in Sub-Saharan Africa

	(1)	(2)	(3)	(4)
Dependent variable	log(publications)	log(publications with OECD coauthors)	log(publications with ex-colonial coauthors)	log(publications in regional journals)
LH_{jt}	0.29***	0.35***	0.35***	0.25***
	(0.018)	(0.025)	(0.023)	(0.024)
LA_{jt}	0.30***	0.12**	-0.057	-0.049
	(0.029)	(0.040)	(0.037)	(0.039)
LS_{-jt} ex-colonial power	0.24***	-0.12*	-0.083	-0.23***
	(0.047)	(0.064)	(0.058)	(0.061)
Controls				
Log population	0.34**	0.98***	1.22***	0.95***
	(0.12)	(0.17)	(0.15)	(0.16)
Year fixed effects	yes	yes	yes	yes
Country fixed effects	yes	yes	yes	yes
R^2	0.95	0.89	0.89	0.87
Adjusted R^2	0.94	0.89	0.88	0.86
Number of observations	1,845	1,845	1,845	1,845

Note: Estimates stem from fixed effects ordinary least square specifications with dependent variables being log of counts of outcomes of country j in year t+2. Robust standard errors are given in parentheses. *, **, *** denote statistical significance at p-values of 0.1, 0.05 and 0.01 respectively.

Table 15: Robustness Checks

Dependent Variable = log(publication count)	(1)	(2)	(3)	(4)
LH_{jt}	0.29***	0.31***	0.24***	0.30***
	(0.018)	(0.019)	(0.024)	(0.021)
LA_{jt}	0.30***	0.29***	0.26***	0.29***
	(0.029)	(0.030)	(0.026)	(0.044)
LS_{-jt} ex-colonial power	0.24***	0.29***	0.25***	
	(0.047)	(0.025)	(0.070)	
LS_{-jt} ex-colonial power × white settler dummy		-0.049*		
		(0.025)		
LS_{-jt} ex-colonial power weighted by imports			0.0013	
			(0.0045)	
LS_{-jt} OECD (excl ex-colonial power) common language				-0.93*
				(0.50)
Controls				
Log population	yes	yes	yes	yes
Year fixed effects	yes	yes	yes	yes
Country fixed effects		yes	yes	yes
R^2	0.95	0.95	0.95	0.94
Adjusted R^2	0.94	0.94	0.94	0.94
Number of observations	1,845	1,845	1,305	1,517

Note: Estimates stem from fixed effects ordinary least square specifications with dependent variables being log of counts of outcomes of country j in year t+2. Robust standard errors are given in parentheses. *, **, *** denote statistical significance at p-values of 0.1, 0.05 and 0.01 respectively.