Partha Ray · Runa Sarkar · Anindya Sen
Editors

Economics, Management and Sustainability

Essays in Honour of Anup Sinha

Springer

Economics, Management and Sustainability

Partha Ray · Runa Sarkar · Anindya Sen
Editors

Economics, Management and Sustainability

Essays in Honour of Anup Sinha

Springer

Editors
Partha Ray
Indian Institute of Management
(IIM) Calcutta
Kolkata, India

Anindya Sen
Indian Institute of Management
(IIM) Calcutta
Kolkata, India

Runa Sarkar
Indian Institute of Management
(IIM) Calcutta
Kolkata, India

ISBN 978-981-13-1893-1 ISBN 978-981-13-1894-8 (eBook)
https://doi.org/10.1007/978-981-13-1894-8

Library of Congress Control Number: 2018950938

This Springer imprint is published by the registered company Springer Nature Singapore Pte Ltd.
The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

Professor Anup K. Sinha

Preface

It is not very often that one comes across an individual who is an outstanding teacher, a well-known researcher, an able academic administrator and a warm, approachable human being. Professor Anup Sinha is one such person.

He was educated in Presidency College, Kolkata; University of Rochester, New York; and University of Southern California, Los Angeles. After completing his Ph.D. from the University of Southern California in 1983, he returned to Presidency College and taught at the Centre for Economic Studies for a number of years. In 1991, he decided to join the Indian Institute of Management Calcutta (IIMC) where he taught till his retirement in 2016. In his long career spanning more than four decades, he has been a visiting faculty in a number of institutions in India and abroad, including Indian Statistical Institute, University of Calcutta, National Institute of Public Finance and Policy, University of Southern California, Washington University in St. Louis, Curtin University of Technology at Perth and Kyoto University. A very popular teacher with an infectious ability to quickly build a strong rapport with his students, it comes as little surprise that he was voted the best faculty for a number of years in IIMC by students and alumni alike.

His doctoral work was on economic development, and he taught the subject over many years as a teacher. He also taught macroeconomics to undergraduate as well as postgraduate students. But his research interest of late went beyond the traditional confines of macroeconomics and economic development. His major publications are in the areas of economic development and reforms, macroeconomic policies, globalization, business ethics and sustainable development. Over the years, he moved away from the mainstream macroeconomics and economic development to the more interdisciplinary area of "sustainable development".[1]

[1] Some of his recent writings bear testimony to this. See, for example, "Corporate Ethics" (in Kaushik Basu (edited) *The Oxford Companion to Economics in India*, Oxford University Press, Delhi, 2007); "The Aura of Green: Commitment in an Age of Uncertainty" (in *Decision* Volume 36, No. 2 August 2009, with Jamie Gilpin); "Good Governance and Sustainability: Making Sense of a Complex Agenda" (in S. Singh-Sengupta (edited) *Spiritual and Ethical Foundations of Organizational Development*, Macmillan Delhi, 2009); "Sustainability: Ethics and the Future" (in *Journal of Human Values*, October 2013); "The Business of Development: A Case Study of Participation and Dependency" (in *Decision* June 2014); and "Sustainable Development and the Concept of a Good Life" (in Runa Sarkar and Annapurna Shaw (eds.): *Essays on Sustainability and Management: Emerging Perspectives*, Springer, 2016).

He has edited three volumes and co-authored two books, one titled *Another Development: Participation, Empowerment and Well-being in Rural India* and the other titled *Economics of Sustainable Development*.[2]

He was active, able and reputed as an academic administrator. He served as Convener of the Faculty Council, Chairperson of the doctoral programme, Editor of IIMC's academic journal DECISION, Dean and three terms as Faculty Representative on the Board of Governors at IIMC. He served on the Board of National Bank for Agricultural and Rural Development (NABARD) from 2006 to 2009. He also serves on the Life Insurance Council of India as an Insurance Regulatory and Development Authority of India (IRDAI) nominee. He is a trustee of Uttaryan, a non-governmental organization which works with mentally challenged children and young adults.

However, the complete gamut of his interests is not restricted to academics alone, but is wide and varied. He learnt to play the violin when in school and played the first violin in the school orchestra. He was a well-known debater during his college days. His passions (according to he himself) lie in the following; cricket, reading crime fiction and listening to classical music—the three "C"s of his life. His interest in theatre is much more than pedantic—he performed in faculty plays at IIMC on two occasions. He also frequently contributes op-ed pieces in newspapers on a variety of contemporary social problems.

In this volume, his colleagues and students have come together to honour him as a memorable teacher, a reliable friend and a wise mentor.

In consonance with his interest, chapters in this volume are arranged in four broad themes, viz. economic development; vulnerabilities and inclusive growth; sustainability and corporate governance; and innovation and management. As a prelude to the volume, what follows below is a brief description of these chapters.

Economic Development

Chapters in this part cover varied themes of economic development such as construction of an adequate indicator of development, issues on governance, political economy and public–private partnership.

Maitreesh Ghatak discusses attempts to develop an adequate indicator of development and in the process highlights the many-sided nature of development. For theorists, the nature of a relevant indicator often depends on a person's view of the goals of development, and hence, there are disagreements about the proper indicator that should be used. It has been known for long that a single-minded focus on per capita GDP does little justice to the idea of development in all its richness. Alternative measures like the percentage of population below the poverty line,

[2] *Another Development: Participation, Empowerment and Well-being in Rural India* (with Runa Sarkar); Routledge, Taylor and Francis, New Delhi, 2015; *Economics of Sustainable Development* (with Runa Sarkar), Business Expert Press, New York, 2018.

measures of inequality such as the Gini coefficient and human development indicators have been proposed. But objective as these may appear, no indicator can truly capture all the dimensions of something as multifaceted as the quality of life. Ghatak notes, for example, that depletion of natural resources and pollution during the growth process and discrimination against the girl child are two important aspects of the development process which are not being addressed via the standard measures. Moreover, development policies cannot be examined in isolation of the political setting. Dictatorships seem to be more efficient in single-mindedly pursuing their visions of development, but due to the absence of the checks and balances (normally present in a democracy), they may not become aware of other deleterious consequences of the path they are pursuing before it is too late.

Sumon Kumar Bhaumik takes up the related issue of governance. For him, governance encompasses both the formal and the informal bases for the relationship between the state and the private citizens as well as the relationship among private citizens. One dimension of governance is "rule of law", and data across countries over time do not suggest that rule of law automatically improves with an increase in per capita income. However, some research seems to indicate that governance quality is higher in democratic countries because in democracies officers are more accountable and there is a higher degree of transparency. Governments can be non-benevolent and rent-seeking, and certain institutions are needed to keep them in check. One cannot rely on the process of growth alone for the emergence of such institutions. Sometimes, external intervention in the form of FDI or aid can be of help, but these can be double-edged. On a more specific note, the economic approach to governance focuses on the mechanisms that facilitate transactions in modern exchange economies. The establishment of clear rules governing transactions must be accompanied by the equal ability of all parties to get the rules implemented. Carrying the idea of a non-benevolent state further, one can envisage situations in which the state itself acts as an intermediary to violate the rules and expropriate some of the parties engaged in transactions. The state can enter into contracts with different groups to enable it to expropriate other groups and then share the revenue with the former. The elites running the state apparatus must co-opt some groups through the distribution of spoils of expropriation and create "patron-client networks that extend down to the rest of the society" for winning the violent confrontations that the expropriation requires. Bhaumik feels that this view of governance indicates that only an "open access order" where political factions have to enjoy the support of social and economic interests, broadly defined, can act as an antidote to this fundamental premise of a non-benevolent state trying to bend contractual rules through violence.

Amitava Krishna Dutt goes back into the history of economic thought and notes that the study of the economy, which used to be called the political economy, changed to economics and the term political economy fell into disfavour. More recently, the term has experienced a revival, although not necessarily referring to the same thing as economics. His chapter describes how the name change occurred, how the term returned and how political economy is distinguished from economics, the change in the nature of the study of economies that accompanied—though not

precisely—the change in the name, the problems caused by these changes and the reasons for the change. He argues that a return to the name political economy from economics and a return to what political economy tried to do before the change in name are desirable not only for a better understanding of the economy but also for the well-being of people, especially those who have been excluded and marginalized.

In contrast to the broad sweep of the other three chapters in Part I, Indrani Roy Chowdhury and Prabal Roy Chowdhury analyse a very specific tool now finding widespread use—the public–private partnership. They analyse the possibility of collusion between the private firm and the government department in the process of PPP formation. They develop a simple model based on risk-sharing to look at this possibility. They show that PPPs are most likely to form in case the externality gains out of the project are significant, and agents are risk-averse. Otherwise, PPP formation may lead to bribery and sub-optimal project choice. In the light of this possibility, the government may opt for direct control over the project instead of forming a PPP.

Vulnerabilities and Inclusive Growth

The chapters in this part are spanned over both theoretical and empirical aspects of vulnerabilities and inclusive growth. These cover theoretical issues as diverse as explanations of the recent Greek crisis, the extent of private contribution to higher education, and the case for and against a unitary education policy across all parts of the society. These apart, there is an empirical and conceptual contribution that looks into the definition of unemployment in India.

Ghosh and Ghosh in their chapter, "Capitalism, Crisis and the Common Man", develop a simple model that redresses many of the major deficiencies of the characterization of the financial sector in the standard IS-LM-based open-economy macro-models. Going beyond Bernanke–Blinder (1988)-type IS-LM model with a credit market, their extended version of an open-economy Keynesian model incorporates financial intermediaries and imperfect capital mobility. In their model, it is shown that the multiplier process that occurs in the real sector and the money or credit multiplier process that occurs in the financial sector take place simultaneously reinforcing each other. They apply this model to explain some of the stylized facts of the Greek crisis. Their results tend to indicate that the higher growth rate in Greece since 1997 could have been due to higher growth rates of GDP in other European countries and USA and the higher growth rate of inflow of capital. Furthermore, the higher growth rates in other European countries and USA and the higher growth rate in net inflows of capital could have brought about sharp falls in interest rates. This could have induced the Greek government to borrow on a large scale to finance additional expenditure which is the primary reason behind the accumulation of a sizable amount of debt by the Greek government by the beginning of 2008.

Dasgupta in his chapter "School Language Policy, Crime and the Minority Underclass" develops a theoretical model of a society consisting of a majority and a minority. These communities differ in terms of a set of behavioural-expressive traits and conventions. Individuals born into a community acquire that community's traits and conventions as part of their upbringing within the community. In such a set-up, he examines the case for linguistic–cultural unification of the educational system in societies with a majority and a minority ethnolinguistic community. It is demonstrated that that possible aggregate efficiency gains from such unification have to be balanced against the consequences of greater income inequality within the minority community. Such expansion may set in motion attempts to expropriate productive individuals which, through cumulative causation, may more than dissipate any income gains accruing to the minority community from integration. Thus, the efficiency case for a unitary education policy needs to be qualified by the possibility of both immiserization and criminalization of the minority.

Bag and Mondol in their chapter, "Private Giving in Higher Education", start with a stylized fact about the significance of private giving in higher education in USA by alumni and top philanthropists as against its non-existence in a country like India, independent of the wealth differential between an American and an Indian. They view education as a consumption good but to be provided only voluntarily. In particular, the greater the collective contributions to education, the better is the quality of institutions where young people can gather knowledge that serves them not only for future careers but also in the enrichment of life experience. Their starting point is Krugman (1979)'s model, wherein it has been shown that in an economy with only private goods, consumers with a preference for product variety, economies of scale in production and monopolistic competition, different regions could have a tendency to merge into a single conglomerate region. Their formal models show that high labour cost due to low population base could have made private good more costly and public good relatively more attractive in an advanced country.

Dutta and Husain in their chapter, "Being Out of Work: An Analysis of Unemployment and Its Duration in India", look into some of the data-related issues on measuring unemployment in India and question the currently used definition of unemployment in India by different agencies like NSSO, Labour Bureau or Census. Using data from the NSSO's 68th round survey, they propose a new definition of unemployment which utilizes information on the duration over which the respondent is without work in the year preceding the survey. They estimate the incidence of such "out of work" respondents and identify groups who are most at risk of being "out of work" using a two-stage least square logistic model. They argue that policymakers and researchers have failed to utilize the potential of such information that not only can generate more realistic levels of unemployment but also can provide information on the duration of unemployment.

Sustainability and Corporate Governance

Chapters in this part look into issues relating to participation for sustainable ecosystems, cases on ecosystem service value and its applicability to business, and corporate governance concerns such as the relationship between ownership and firm performance in India.

Banerjee in her chapter, "Sustainable Eco-Management: Participatory Mechanisms and Institutions", draws on a number of studies conducted between 2007 and 2014 on different rural, peri-urban and urban pockets of the state of West Bengal, India, focusing on ecologically sustainable management of natural resource and environments like social forestry, wetland fisheries and municipal solid waste disposal in a co-management framework with active beneficiary participation. The objectives of these studies were to explore the context specificity of the success probability of co-management practices in different situations. Specifically, two different situations where otherwise suitable projects for participatory resource management failed to attain the intended result due to some peculiarities of the cases related to the presence of some built-in contradictions are studied. In one situation, the composition of the stakeholder group led to a deviation in the equal participation norms, and in the other, the regulatory set-up comprised multiple authorities with inherent jurisdictional conflicts. Banerjee's assessment validates the prevalent understanding among social scientists that though the participatory approach is conceptually more democratic, its success potential is highly dependent on the local conditions.

Ghosh introduces the notion of creating share value while discussing the importance of adoption of sustainability as a corporate strategy by businesses in the developing world. He views sustainability primarily from the perspective of biodiversity conservation. Businesses inherently depend on the ecosystem services, that is, services provided for free by the ecosystem to the human community. Not acknowledging the value of these services in the course of the working of the firm may affect the firm's long-running bottom lines. The firm's ecological footprint that comes in the way of biodiversity conservation and degrades ecological health essentially erodes the "natural capital" of the planet on which the firm is dependent. The importance of ecosystem services and their valuation in shared value creation are paramount. It then emerges that by embracing sustainability as a corporate strategy, firms are essentially creating shared value. This is exhibited using two cases of ecosystem service values at two different scales: one at the scale of a wetland ecosystem, the Kunnigal Wetland, and the other at the scale of a landscape, the Terai Arc Landscape. The notion of ecosystem services as "GDP of the poor" thereby linking ecosystem services to livelihoods is also introduced.

The relationship between ownership and firm performance has been an area of intense debate in the corporate governance literature. There are two competing hypotheses in the existing theoretical literature regarding the effects of insider ownership on firm performance. The first one argues that an increasing insider ownership aligns the manager's interests with outside shareholders' and hence

results in a positive effect on firm performance. The second diametrically opposite view is the entrenchment hypothesis. It suggests that since higher insider managerial shareholdings are likely to shelter insiders from the influence of market for corporate control, firm performance is adversely affected. Chakraborty examines the relationship between insider ownership and firm performance in Indian listed firms using a panel semi-parametric regression technique. A new structure to the insider ownership–performance relationship which captures a more complex characterization of the evolving behaviour of managers in Indian firms dominated by business groups is proposed. The results establish that the relationship is quartic for Tobin's q. Further, she argues that with equity holdings above 50%, although the managers have substantial control of the firm, the internal governance resulting from the corporate governance mechanism will lead to convergence of interests. Only at very high levels of managerial holdings—above 80% of equity holding—the entrenchment effect predominates.

Innovation and Management

Under the broad theme of innovation and management, chapters in this part are centred around the themes of business incubation, the role played by the human brain in economic decision-making by drawing upon the current research on neuroeconomics, the impact of information and communications technology (ICT) on a business manager's life and leadership role.

In underlining the role of business incubation in promoting new business enterprises, Bhaskar and Phani propose a generic framework of a business incubator model for a sustainable innovation ecosystem. The authors start by impressing upon the reader that the extent to which new business enterprises contribute to a nation's economic growth depends on the optimality of the utilization of its resources. This, in turn, can be achieved by increasing the efficiency of the existing business enterprises and the promotion of new ones. This not only justifies the use of protectionist policy measures such as subsidies, quotas and tariffs, but also underlines the need for more progressive measures such as promoting innovation, developing S&T infrastructure and supporting business incubation services. A critical review of the available literature points to the criticality of the availability of knowledge, finance and other crucial resources to new business enterprises, but there is no convergence on what constitutes an effective or ineffective business incubation process. Bhaskar and Phani advocate developing a framework for delivery and impact assessment of business incubation services. Innovation policy should be designed to facilitate a robust business incubator business enterprise engagement network to leverage their individual strengths across this network to accelerate and sustain the growth of innovation ecosystem.

The next chapter by Sharda examines the role played by the human brain in economic decision-making by synthesizing the current research on neuroeconomics. Neuroeconomics combines classical economics and neuroscience to

deepen our understanding of the role played by the human brain in economic decision-making. It provides a mechanistic, mathematical and behavioural framework to understand choice-making behaviour. Sharda identifies major themes and trends in the literature and presents potential areas of research related to the neurobiology of economic decision-making. She explores the role of reinforcement learning systems in valuation and choice, value-based decision-making and decision-making under conditions of risk and ambiguity. Further, social preferences and context dependencies in decision-making are also explored. The chapter ends with some thoughts on the fallacies in interpretations, methodological issues in neuroeconomics research, current concerns and future directions.

Continuing with the emphasis on understanding human organizations, the impact of all-pervasive information and communications technology (ICT) on a manager's life is explored next. Notwithstanding their benefits ranging from improved efficiency, flexibility and social connectivity, there are some overwhelmingly disturbing impacts that are less explored. In their chapter on the dark side effects of ICT, Tarafdar and Stich examine technostress, technology addiction and information overload as negative externalities arising out of the increasingly pervasive use of ICT. The authors examine the negative consequences of these outcomes and identify possible mitigation mechanisms.

The human element in business is explored further in the last chapter of this part (and book), which examines what it takes to be a good leader. As an executive coach who comes across people from different walks of life, Chatterjee elaborates on the qualities of leadership and how they can be illustrated through the life of Anup Sinha.

These chapters are contributions from Anup's erstwhile students, colleagues and friends who consider Anup a great academic and one of the finest human beings they have come across; someone who did not keep his thinking and communication in the narrow confines of mainstream economics, was never shy of newer areas, and probed into issues relating to sustainability across generations.

Kolkata, India

Partha Ray
Runa Sarkar
Anindya Sen

Contents

Editors and Contributors

About the Editors

Partha Ray is currently Professor of economics at the Indian Institute of Management Calcutta. From 2007 to 2011, he was Adviser to Executive Director (India) at the International Monetary Fund, Washington, D.C. Earlier, he was working in the specialist cadre of Economists in Reserve Bank of India's Economic Research Department for nearly 16 years in various capacities; his last position was Director, Financial Markets Division. Educated in Calcutta, Mumbai and Oxford, he has published extensively on macroeconomics, monetary policy and global financial crisis. His recent publications include *Financial and Fiscal Policies: Crises and New Realities* (with Y. V. Reddy and Narayan Valluri, 2014); *Monetary Policy: Oxford India Short Introduction* (2013); and *The Global Economic Crisis through an Indian Looking Glass* (with A. Kishore and M. D. Patra, 2011).

Runa Sarkar is Professor of economics and Dean (Academic) at the Indian Institute of Management Calcutta. Prior to this, she taught at the Indian Institute of Technology Kanpur. She, along with Anup Sinha, is the co-author of two books, one titled *Another Development: Participation, Empowerment and Well-being in Rural India* and the other titled *Economics of Sustainable Development*. She is one of the co-editors of the *India Infrastructure Report,* 2010 and 2009. Her latest book titled *Environment, Business, Institutions* was published in 2017. She has also co-edited another book, titled *Essays on Sustainability and Management: Emerging Perspectives*. Her areas of interest are corporate sustainability and the application of social informatics in agriculture. She is currently Chairperson of CTRAN Consulting and is on the board of two other companies of the BASIX Group.

Anindya Sen is Professor of economics at the Indian Institute of Management Calcutta. He was earlier Visiting Professor at School of Business and Management, American University of Armenia, and at the Department of Economics, University of Notre Dame, USA. He was also Professor at Indira Gandhi Institute of

Development Research, Mumbai. He has earlier held the positions of Dean (Programme Initiatives) and Dean (Academic) at IIMC and Dean at IGIDR, Mumbai. He was Director Incharge of IIM Ranchi from November 2014 to February 2017. He holds a Ph.D. from the University of Southern California, USA. His fields of interest are industrial organization, theory of the firm, regulation, law and economics and strategic management. He has authored and edited several volumes. He has been the general editor for a cluster of books on economics and development in the *Oxford India Short Introduction series*. Currently, he is editing two series for Routledge. He has published in both Indian and international journals and contributed chapters to a number of volumes.

Contributors

Parimal Bag Faculty of Arts and Social Sciences, Department of Economics, National University of Singapore, Singapore, Singapore

Sarmila Banerjee University of Calcutta, Kolkata, India

R. A. Bhaskar Indian Institute of Technology (IIT) Kanpur, Kanpur, India

Sumon Kumar Bhaumik Sheffield University Management School, University of Sheffield, Sheffield, UK; IZA—Institute of Labor Economics, Bonn, Germany; Global Labor Organization, Geneva, Switzerland

Indrani Chakraborty Institute of Development Studies Kolkata, Bidhan Nagar, India

Anamitra Chatterjee Ashridge Executive Education Hult, Berkhamsted, UK

Indrani Roy Chowdhury Jawaharlal Nehru University, New Delhi, India

Prabal Roy Chowdhury Indian Statistical Institute, Delhi Center, New Delhi, India

Indraneel Dasgupta Economic Research Unit, Indian Statistical Institute, Kolkata, West Bengal, India

Amitava Krishna Dutt Department of Political Science, University of Notre Dame, Notre Dame, USA; FLACSO, Quito, Ecuador

Mousumi Dutta Economics Department, Presidency University, Kolkata, India

Maitreesh Ghatak London School of Economics, London, UK

Ambar Ghosh Jadavpur University, Kolkata, India

Chandana Ghosh Indian Statistical Institute Calcutta, Kolkata, India

Nilanjan Ghosh World Wide Fund for Nature, New Delhi, India; Department of Economics, Observer Research Foundation, Kolkata, India

Zakir Husain Humanities & Social Sciences Department, Indian Institute of Technology, Kharagpur, India

Debasis Mondal Department of Humanities and Social Sciences, Indian Institute of Technology Delhi, New Delhi, India

B. V. Phani Indian Institute of Technology (IIT) Kanpur, Kanpur, India

Kirti Sharda Indian Institute of Management Ahmedabad, Ahmedabad, India

Jean-François Stich ICN Business School, Nancy, France

Monideepa Tarafdar Lancaster University (Management School), Lancaster, UK

Part I
Economic Development

Measures of Development—Concepts, Causality, and Context

Maitreesh Ghatak

1 Alternative Measures of Development

Behind the very concept of development there lie two things: first, a gap between the actual and the possible; and second, a hope that a certain process of change will translate the potential into reality.

However, beyond this basic conceptual point, there is very little that people universally agree on regarding development. Just to give two prominent examples, consider the notions proposed by Robert Lucas and Amartya Sen. To Lucas (1988) the field of development economics involves studying what explains the variation over time and across individuals, households, regions, and countries of per capita income. In contrast, Sen (1988, 1999) argues that human development is about the expansion of capabilities of individuals, by increasing access and opportunities to the things they have reason to value. There has been a significant amount of work in the last few decades in development economics that has advanced our understanding of what constraints the economic potential of the poor, whatever is our notion of development (see, for example, Banerjee and Duflo 2010; Ghatak 2015a).

Some of the disagreement has to do with the goals of development which in turn depends on one's ideal view of society and quality of life for an individual. This is often expressed as a disagreement over the correct measure or index of development. Some of the disagreement has to do with what is the right path to achieve a certain goal, even when there is no disagreement over objectives.

The goals of development inevitably involve some subjective judgement, but there are a number of objective indicators that are popularly used to measure different aspects of development. Some popular examples of such indices would include per capita income, percentage of the population below the poverty line, measures of inequality such as the Gini coefficient, and human development indicators. But

M. Ghatak (✉)
London School of Economics, Houghton Street, London, UK
e-mail: m.ghatak@lse.ac.uk

P. Ray et al. (eds.), *Economics, Management and Sustainability*,
https://doi.org/10.1007/978-981-13-1894-8_1

objective as these may appear, no indicator can truly capture all the dimensions of a something as multifaceted as the quality of life. Which indicator do we use to determine measure development, then? And when each of these indicators throws up a different image, how are we to arrive at a more or less accurate overall picture?

The simplest among the development indicators is per capita income and the rate of its growth. Despite its popularity, however, per capita income has quite a few limitations as an index of development. For one, it can capture the value of only those goods and services that are bought and sold in the market. The actual elements that determine one's quality of life, such as education, health, environment, infrastructure, and law and order, remain outside its ambit. As do indicators such as life satisfaction and happiness, which are translated into numbers these days, based on answers from respondents to questions like 'Would you describe yourself as happy?' Economists such as Richard Layard (see, for example, Layard 2006) hold this to be the best indicator of development since the rest are merely inputs contributing to it. However, this indicator has no objective yardstick and depends entirely on individual perception and social influence. That is why it is not of much help when comparing two countries or two different periods of time.

If per capita income is too narrow an index of development, and life satisfaction too diffuse, then perhaps a middle ground could be reached by adopting the human development index. At the basis of this index lies Amartya Sen's capability theory. According to this theory, the goal of development is the gradual enhancement of an individual's capability; while an individual's well-being cannot be determined by policies or cardinally measured, it can be safely said that enhancing her capability will enable her to realize her goals. This index formulated has been appearing in the UN Development Report since 1990, and it comes up inevitably in any discussion on development (see, Anand and Sen 1994). It is essentially the average of three different indices—per capita income, mortality rate which is the index of health, and an index of education. The third used to be determined earlier by the rate of adult literacy and enrolment rate in schools, but from 2010, the basis shifted to mean years of schooling.

Education and health do not merely contribute to the rise of national income, but they are also important indicators of our quality of life. While Flaubert may have been laughing at our 'pursuit of happiness' when he laid down 'the three requirements of happiness' ('To be stupid, selfish, and have good health are three requirements for happiness, though if stupidity is lacking, all is lost'), he could not ignore the importance of good health. But capability cannot be determined by education and health alone; it depends considerably on the rights that citizens are allowed to enjoy. For instance, education and health indicators can hardly paint the true picture of development in countries where a citizen's democratic rights are violated regularly. Such violations can range from curbing one's freedom of expression, to giving a free rein to crime, extortion, and violence, using the law and order machinery to commit extra-legal atrocities, and taking away people's property and livelihoods (often in the name of development). Some organizations, such as Amnesty International, have been publishing indices based on citizens' rights, but development indices have not reflected this aspect so far.

That's just one side of the story. The per capita income approach also fails to capture inequality among sections of the population. In other words, it is quite possible for two countries—one with abject poverty and abundant affluence existing side by side and the other with a more equitable distribution of wealth—to have the same per capita income. If our democratic principles mandate that every individual's well-being count in the measure of social welfare, then our development index must dig deeper than collective income and take into account the distribution of wealth.

Most economists work around this problem by looking at the percentage of the population below the poverty line. But this method offers only a limited perspective on the problem of inequality as inequalities may exist among those who live above the poverty line, as well as between those living above and below the poverty line. Moreover, we need to keep in mind that inequality is a relative indicator while the poverty line is an absolute one, so the two need not necessarily be connected, and one may well be greater or lesser than the other.

2 Problems with the Standard Measures of Development

We have talked so far about the four main indices to 'measure' development, viz. per capita income, human development, percentage of the population below the poverty line, and the Gini coefficient. There are a few other indicators which get mentioned in debates on development, and most of these serve as alternatives to the conventionally used human development indicators; these include infant mortality rate, measures of child malnutrition, school enrolment rates, and school completion rates. Dreze and Sen (2013) provide several examples of such indicators. However, there are yet more measures of development that are qualitatively different from the ones we just discussed. We cannot possibly engage with all of them in the space of this article, but it is important to mention, however, briefly, a few of these measures.

First, when we calculate national income, we must allow for the depreciation of the capital stock needed to generate it. The idea behind this is that national income is not a one-time product, and the implicit assumption is that the economy is capable of continuing to produce the same or higher levels of national income. However, this mode of calculation has no room for the devaluation or depletion of natural resources. But we happen to live amidst ample proof of pollution and abuse of our natural environment all around us. Unlike buildings, infrastructure and machinery, natural resources cannot be rebuilt or replenished when needed. Changes in climate are taking a heavy toll on our farmers, pollution is leaving a few more of us sick every day, and unbridled construction work is leading to natural disasters and destruction all around. What makes the grim picture truly scary is that the price for our greed and downright apathy will have to be paid by the generations to come. Sadly, economists are yet to take cognizance of this threat. Only a few exceptions such as Arrow et al. (2004) have argued in favour of including natural wealth and quality of environment, along with income and human resources, in formulations of development index. At an outside guess, one could say that China or India's development story would not

appear so impressive if one took into account the environmental impact along with income growth.

Second, a society where girl children are subject to the worst kind of discrimination cannot be evaluated with an overall development index that does not take this aspect into account. In education and health, for instance, girls often lag far behind boys, and so we need to apply the indices for these two parameters to girls and boys separately. In discussing development, we economists tend to assume that the causes of inequality are economic. But not all the factors determining one's economic status are economic in nature; some are rooted in social conditions. When girls and women are discriminated against in educational institutions and the workplace, we cannot possibly use the correlation between capability and economic status to arrive at any conclusion about overall efficiency of using talent and skills in the economy. The same goes for other forms of social discrimination, such as the kind faced by lower-caste groups or ethnic minorities. Even if we were to ignore the ethical aspects of discrimination, we cannot deny the adverse effect that such forms of discrimination have on overall economic efficiency of a country. This would create an extra gap between the potential and the reality in terms of development indicators such as per capita income. Therefore, no matter what overall or aggregative index we use to measure development, we must take into account the relative development and growth rate across genders and social groups along with conventional indicators such as per capita income and rate of economic growth to capture social inequality.

Third, even if we are to concentrate solely on the economic aspect of development, we still come up against one important stumbling block—the indicators do not capture fluctuations in income patterns at all. Per capita income may be above the poverty line, but the more this income is liable to uncertainty, the lower should be the value of the development index. Theoretically, it seems eminently possible to formulate such an index, but in actual practice, we hardly ever come across one.

In short, every index of development has its particular strengths and limitations. Just as we can conduct a battery of tests on the human body and diagnose different ailments by analysing the results, so can we diagnose various development-related maladies from what the different indicators tell us. An added complication relative to the analogy with the human body is the problem of aggregation—the development of a country depends on the development of its individual citizens. Which aspects of a citizen's development should count, whether they should be objective or subjective, and how the data from all these indicators could be synthesized into an overall development index is a conceptually complex exercise. If the citizens of a country or state were all equal in all respects, then their per capita income and life span would be the same and their average would give us a fair and square development index without causing added concerns about inequality. But in such a scenario too, opinions might differ on which of these is the best measure of quality of life—income, life span, education, health, or life satisfaction. And then, even if we agree on one (or more) of these indicators as the best, we would still be fighting over the right way to measure inequality, because, after all, no two human beings are equal.

3 Symptoms Versus Causes of Underdevelopment

Economists who put per capita income and the rate of its growth above all else usually fall into two groups. The first group considers these to be the best indicators of development, while the second group feels that improvement on these counts is the best way to pull up the others (for instance, poverty alleviation). The first group sees income disparity not as a problem but as a natural outcome of the fact that some people are more capable and hard-working than others. They are also opposed to the idea of redistributing wealth to reduce inequality, on the ground that the step will raise the tax burden on citizens, leading to a fall in productivity as well as in investment, which in turn will culminate in a diminished national income. Strange as it may sound, this view is quite popular among Western right-wing economists, but has few proponents in India. So by Western standards, even our right-wing economists would appear to be quite left-leaning, given that they do not deny the importance of poverty alleviation as a social goal, whatever be their views on the means to that end. This is probably due to the fact that in the Indian context, it is impossible to deny that those who perform the hardest physical labour are poor, or that poverty is most often the result of a lack of opportunity. So the advocates of the growth-of-per-capita-income route to development see it as the best way of creating opportunity and employment for the poor, and of increasing government revenue, which could then be channelled into poverty alleviation programmes.

On the other hand, those who prioritize inequality in development calculations feel that the development of human capital of the poor would push up national income in the long run by expanding its human capital base, and hence should be considered as an investment. In their view, therefore, redistribution, if kept within a limit, is a positive thing. The difference between these two groups of economists is thus not always one of ideology, but over the best means to the goal of development—just as the disagreement between the Amartya Sen and Jagdish Bhagwati camps in the context of development policy in India is not over the need to remove poverty, but over the road taken to attain that goal.

But the indicators of development alone are not enough to provide a reliable roadmap. As mentioned earlier, these are mere symptoms. The real malaise lies deeper within and needs further investigation. For instance, many are of the opinion that a rise in per capita income brings down poverty, and a cursory glance at the statistics will, in fact, throw up a correlation between the two indicators. But which among the two is the cause, and which the effect? It seems possible enough that if per capita income rises, then poverty could fall, but it seems equally plausible that if poverty falls, per capita income would rise. Or it could just as well be that some other extraneous factor such as a shift in government policy or a change in the economic environment (such as investment in infrastructure, use of new technology in farming, opening up of new export markets, or expansion of the banking system) is responsible for a rise in per capita income and a fall in poverty levels.

To understand the effect of government policies and the economic environment on the different development indicators, we need to go beyond exploring correlations and

establish causality instead. Mainstream development economics has of late been more mindful of this need and recognizes that suggesting outlay amounts is not enough since the lion's share of the allocated amount is wasted or siphoned away before it can reach the poor. The emphasis now is on giving concrete recommendations on planning and implementation of new schemes on education, health, microfinance, farm technology, and poverty alleviation.

4 Development Measures and the Political Context

Let us now turn our attention to the political context of development. Development policies cannot be examined in isolation of the political setting. When a government decides to follow a particular development policy, it is prompted as much, if not more, by political calculations as by ideologies and development indices. The presence of opposition parties, public opinion, the media, the legal system, and civil society in a democracy ensures that the party in power must compromise, or take one step back for every two steps forward in trying to implement its chosen policies. The downside of having so many stakeholders is that often, implementation of policies and schemes that would clearly benefit the most, gets stalled. But thankfully, there is an upside as well.

Notwithstanding what the ruling dispensation's favourite development indicator is, it cannot get complacent with improving performance on that front alone; it has to look at the others as well, or be ready to face uncomfortable questions. An authoritarian political system, however, has no problems of this kind and can carry out unpopular but necessary policies on a short-term basis. But of course, authoritarian regimes are quite likely to act to maximize their own narrow objectives and turn a blind eye to overall indicators of development.

To give an example, raising per capita income is given such enormous importance in China that the fates of provincial administrators hang in fine balance depending on their state's performance on that front. A recent study reveals that in this bid to raise per capita income, most of the provinces have compromised on pollution control (see Jia 2017). In the absence of the traditional checks and balances of a democracy, the pollution levels could actually reach disastrous proportions unless the Central Committee takes notice of the ticking time bomb. China's one-child policy, which has now been partially reversed, too has precipitated its own set of problems. The ratio of women in the population has been steadily going down, leading to the usual social problems.

India too has its share of problems due to pollution or gender discrimination. And of course, there's no denying that China is far ahead of India, not only in levels and growth rate of per capita income, but also in poverty reduction, education, health, and most of the important indicators of development. What is significant, however, is that the two development indicators where China has not managed to beat India fair and square (leaving aside measures of freedom and democracy) are pollution

indicators and the gender ratio. And this, I would like to argue, has a lot to do with the political systems of the two countries.

On the other hand, it is true that if we look solely at the gender-based development indicators, India lags behind Bangladesh, a country that ranks far behind it in terms of national income and its rate of growth. This would invariably raise questions about the importance India's policymakers give to half of the country's population. If one of the fundamental principles of representational democracy is to aim its governance engines at the welfare of the maximum, then we have to admit that India is in the process of development at best, and nowhere close to being an ideal democracy.

The more evolved and developed a democracy, the less risky its efforts at raising the national income. This is because a democratic system's in-built checks and balances (including the media, civil society, and the competition among political parties to come to power) will ensure that the ruling party or coalition cannot ignore the other development indicators and put all its energies behind the growth of national income. Research shows that as a country moves up the democratic scale, expenditure on education and health increases in proportion to national income. Also, the affluence resulting from the growth of national income begins to get reflected in national revenue, raising hopes of a possible increase in spending on social and human development. In countries that show little signs of progress on the democracy front, we have no option but to depend more heavily on the human and social development, and environment-related indicators while determining their performance on the development index. In other words, a country's political status as a democracy or its progress on the democratic scale plays a big role in determining the extent of influence its national income or the rate of growth of per capita income will have on its overall development.

5 Conclusions

How do alternative concepts of development affect our policy debates? At least in the context of India, it seems that economic policy debates are forever stuck in a rather tedious back and forth between two dominant narratives on the state of the economy.[1] One centres around growth (see, for example, Bhagwati and Panagariya 2012) and the other around poverty (see, for example, Dreze and Sen 2013). Anytime one camp talks of double-digit growth and catching up with China, the other points out that India fares worse than sub-Saharan Africa or a relatively poor neighbouring country like Bangladesh in certain social indicators. Every upbeat story about India's growing economic clout seems to be inevitably accompanied by an account bemoaning the abysmal state of India's social indicators.

The problem with the growth-based narrative is that, while growth is necessary for poverty alleviation or improvements in social indicators, it is not sufficient. For example, take the dream growth rate of 10%. It will take 26 years of sustained

[1] This section is based on Ghatak (2015b).

growth of 10% per year in incomes (no country in history has had a quarter century of sustained double-digit growth!) to bring an Indian who is right on the poverty line up to merely the current level of per capita income, which is low by global standards to start with. Growth can bring improvements in standard of living, but a poor person would have to wait for a quarter of a century for even a glimmer of that!

The problem with the redistribution-based narrative is that, if you focus just on redistribution, it would hardly make a dent on poverty. Yes, as attractive it may sound to some to tax the rich, if we take the ratio of total billionaire wealth to GDP, it stood at only 10% in 2012 (starting with 1% in the mid-1990s). If we took all of this wealth and divided it among the poor (350 million Indians), each will get roughly the same as the amount marking the current poverty line (roughly $450 per year), and more importantly, this will be a one-time affair!

Growth is indeed necessary for long-term poverty alleviation. But to take advantage of growth opportunities, the poor need access to human capital, the key inputs to which are education and health. While it is true that in the post-liberalization era growth has indeed lifted millions out of poverty, it is also true that the extent to which growth has made a dent on poverty (growth elasticity of poverty being the technical term) has been lower in India than in China and other comparable countries. This is mainly because of the shockingly low levels of human capital for a large chunk of the population.

Whatever is one's favourite measure of development, the key to sustained increases in standard of living is clearly lies in fostering mobility through investments in human capital. Markets create opportunities for those with human capital, and it is the responsibility of the government to ensure that the poor acquire the human capital necessary to take advantage of these opportunities. Fostering investment in the human capital of children is therefore a win–win strategy—it helps achieve both higher growth rates and reduces poverty, and removes the apparent tension between these objectives as implied by the growth versus redistribution debate. And once people are educated and achieve a certain minimum threshold level of income, they can decide for themselves what is the best path of development suited to them.

Acknowledgements The author remembers with gratitude Prof. Sinha's lectures on development and trade; among the first he was exposed to as an undergraduate student at Presidency College, and many discussions in and outside of the classroom.

References

Anand, S., & Sen, A. (1994). *Human development index*: *Methodology and measurement.* New York: Human Development Report Office Occasional Paper 12, 1994. Reprinted from *Readings in human development,* by S. Fukuda-Parr & A. K. Shiva Kumar, Eds., 2003, New Delhi: Oxford University Press.

Arrow, K., Dasgupta, P., Goulder, L., Daily, G., Ehrlich, P., Heal, G., et al. (2004). Are we consuming too much? *The Journal of Economic Perspectives, 18*(3), 147–172.

Banerjee, A., & Duflo, E. (2010). *Poor economics: Rethinking poverty and the ways to end it*. Noida: Random House.

Bhagwati, J., & Panagariya, A. (2012). *India's tryst with destiny: Debunking myths that undermine progress and addressing new challenges*. Noida: Collins Business.

Dreze, J., & Sen, A. (2013). *An uncertain glory: India and its contradictions*. London: Allen Lane.

Ghatak, M. (2015a). Theories of poverty traps and anti-poverty policies. *World Bank Economic Review, 29*(*Supplement 1*), S77–S105. (Papers and Proceedings of the ABCDE).

Ghatak, M. (2015b). It is time to move beyond the growth versus redistribution debate and focus on economic mobility. *South Asia @ LSE Blog*, http://blogs.lse.ac.uk/southasia/2015/07/22/it-is-time-to-move-beyond-the-growth-vs-redistribution-debate-and-to-focus-on-economic-mobility/.

Jia, R. (2017). *Pollution for promotion*. Working Paper, University of California, San Diego.

Layard, R. (2006). Happiness and public policy: A challenge to the profession. *The Economic Journal, 116*(510), C24–C33.

Lucas, R. E. (1988). On the mechanics of economic development. *Journal of Monetary Economics, 22,* 3–42.

Sen, A. (1988). The concept of development, Chapter 1. In H. Chenery & T. N. Srinivasan. Elsevier Science Publishers B.V.

Sen, A. (1999). *Development as freedom*. Oxford: Oxford University Press.

Governance: Some Observations

Sumon Kumar Bhaumik

1 Introduction

Dixit (2009) noted that "the concept of 'governance' has risen from obscurity to buzzword status in just three decades" (p. 5), at least as measured by the manifold increase in its occurrence in academic archives such as EconLit, between the 1970s and 2008. However, much of this interest in governance has been restricted to the corporate sphere, with authors examining in detail the nature and impact of corporate governance in different contexts. Over the same period of time, the economics literature has, instead, focused on the importance of institutions (North 1987, 1989; Knack and Keefer 1995; Rodrik et al. 2004). Not surprisingly, therefore, the economic approach to governance emphasizes protection of property rights, ease of contract enforcement and provision of public goods through collective action (Dixit 2009). It is, however, well understood that sometimes, especially in the context of developing economies, governance and the government are not synonymous, and that informal institutions such as norms might secure the basis for property rights, contract enforcement and collective action.

The working definition of governance encompasses much more than this narrow focus on the determinants of transaction costs and provision of public goods in an economy. According to UNESCAP, governance "is the process of decision-making

S. K. Bhaumik (✉)
Sheffield University Management School, University of Sheffield, Sheffield, UK
e-mail: s.k.bhaumik@sheffield.ac.uk
URL: https://glabor.org/wp/

S. K. Bhaumik
IZA - Institute of Labor Economics, Bonn, Germany

S. K. Bhaumik
Global Labor Organization, Geneva, Switzerland

P. Ray et al. (eds.), *Economics, Management and Sustainability*,
https://doi.org/10.1007/978-981-13-1894-8_2

and the process by which decisions are implemented (or not implemented)."[1] In other words, it is as much about a government's (arguably laudable) decision to invest in physical infrastructure as its inaction in the face of rampant corruption and subversion of the state machinery. Similarly, it is as much about formulating policies that are aimed at reducing poverty and inequality, as it is about the inclusiveness of these policies irrespective of caste, gender, race, religion and any other socio-demographic marker. Indeed, UNESCAP identifies eight different dimensions of governance, namely participation, rule of law, transparency, responsiveness, consensus orientation, equity and inclusiveness, effectiveness and efficiency, and accountability. In other words, governance encompasses both—what North et al. (2009) call—public and private orders, respectively, the (generally formalized) basis for the relationship between the state and private citizens and the (formal and informal) basis for the relationship among private citizens.

The World Governance Indicators, based on the research of Kaufmann et al. (2010) and available from the World Bank, provide measures of six dimensions of governance across countries, regions and income groups.[2] The percentile ranks for five different income groups, for all six dimensions, are reported in Fig. 1. The figure reports the percentile ranks for 1996 and 2015, spanning a 20-year period which is arguably sufficiently long to change the formal drivers of governance quality. Figure 1 suggests that there is a strong correlation between governance quality and a country's per capita income level, but one cannot infer causality from this figure. The question, therefore, is whether the causality runs from governance quality to income level, i.e., whether governance quality can be changed in a way that can facilitate development that is, by construction, correlated with the income level.

In Fig. 2, we compare the relative strength of one dimension of governance, namely rule of law, across countries, between 1996 and 2015. During these twenty years, the income levels of many countries had changed dramatically. For example, between 1996 and 2015, China's PPP-adjusted per capita GDP (in constant 2011 USD) rose sharply from 2.789 to 13,570, and corresponding figures for India were 2,149 and 5,754, respectively. The percentage growth of per capita income levels in these countries were much more impressive than the corresponding growth in, for example, the UK (from 29,267 to 38,509) and the USA (from 40,501 to 52,790).[3] Fig. 2, however, suggests that the rapid growth in per capita income in countries such as China and India has taken place without much change in their relative positions in the world, at least with respect to rule of law. The change in the absolute measure of rule of law over this 20-year period was small as well. China's measure of rule of law changed from −0.43 in 1996 to −0.34 in 2015, and the corresponding numbers for India were 0.26 to −0.06. By comparison, the measures of rule of law for the UK and the USA were, respectively, 1.59 and 1.45 for 1996, and 1.81 and 1.60 for 2015. In other words, the quality of rule of law in China and India were not only stagnant

[1] Source: http://www.unescap.org/sites/default/files/good-governance.pdf. For other definitions of governance, see Table 1 in Grindle (2007).

[2] For a critique of measures of governance, see Fukuyama (2013).

[3] Source: World Bank Open Data, available from https://data.worldbank.org/.

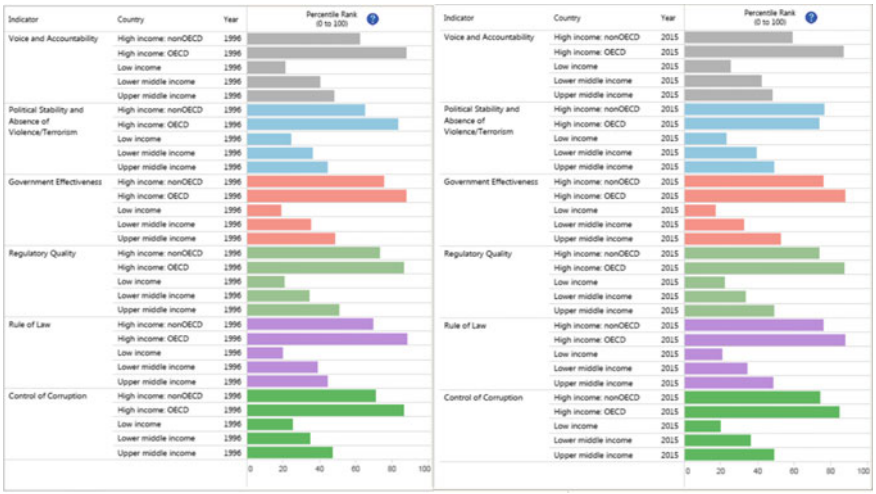


Fig. 1 Governance quality by income group. *Note* The graphs are generated by the interactive tool available on the World Governance Indicators' Web page on the World Bank's Web site: http://info.worldbank.org/governance/wgi/index.aspx#reports. The graph reports the average percentile rank of countries within each income group. The indicators themselves, which are not reported, are measured on a scale of −2.5 to +2.5. Details about the measures can be found in Kaufmann et al. (2010)

relative to other countries, they are also stagnant—indeed, mildly deteriorating—in absolute terms.

This raises the question as to whether or not good governance is a necessary condition to facilitate economic growth. Indeed, it has been argued that "not all governance deficits need to (or can) be tackled at once" and that we may want to instead focus on "the *minimal* conditions of governance necessary to allow political and economic development to occur" (Grindle 2007: p. 554). Similarly, Kurtz and Schrank (2007) argue that even if the institutional changes that are required for good governances are implanted through legislative and judicial action, they may be ineffective until a threshold of economic development is attained in the first place.[4] In other words, any causal relationship between good governance and economic growth may not be as meaningful as a discussion about what drives (or influences) governance and whether it can be altered by way of external interventions. In this paper, we address this latter issue and outline a framework that can be used to understand how governance quality evolves over time in a specific context.

[4]Note also that institutional changes brought about by top-down legislative and judicial measures may not be effective if they lack legitimacy among the citizens. For example, Berkowitz et al. (2003) argue that laws transplanted from other contexts are not effective unless the population of the transplanting country is familiar with the laws already or unless the laws adapted internally. The importance of taking context-specific factors into consideration has also been emphasized by Dixit (2009).

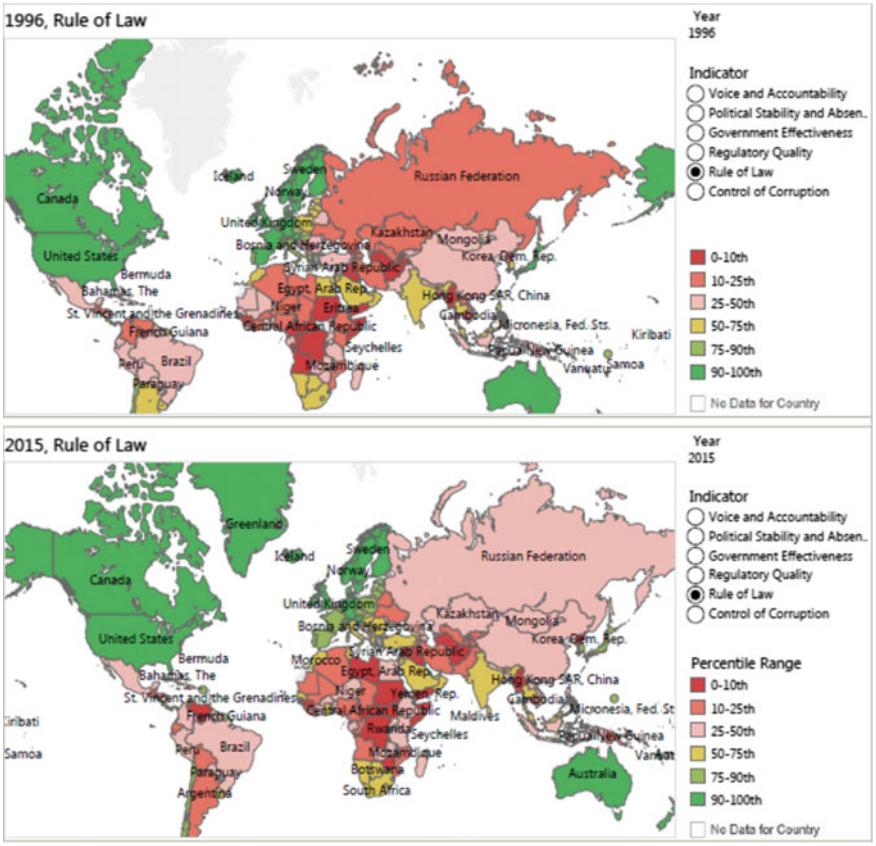


Fig. 2 Rule of law across countries (*Note* The graphs are generated by the interactive tool available on the World Governance Indicators' Web page on the World Bank's Web site: http://info.worldbank.org/governance/wgi/index.aspx#reports. The graph reports relative quality of rule of law across countries. The rule-of-law indicators themselves, which are not reported, are measured on a scale of −2.5 to +2.5. Details about the measure can be found in Kaufmann et al. (2010))

2 Governance: Characterization and External Influence

It is relatively easy to identify the manifestations of weak governance: weak property rights protection, weak rule of law, high levels of corruption, weak accountability of the government, to name a few. Researchers have, in particular, focused on three different manifestations of weak governance, namely democracy, property rights and corruption. In some cases, the political elite directly expropriate the citizens and sometimes the former expropriate the latter on behalf of a small minority of economic elites (Ndulu and O'Connell 1999). Not surprisingly, therefore, the academic literature on governance has focused significantly on the relationship between democracy

and governance.[5] For example, Rivera-Batiz (2002) demonstrates that governance quality is higher in democratic countries, after controlling for other factors, and that democracy leads to better governance by making the officials more accountable and thereby limiting their ability to engage in corruption. In the same vein, Abed and Gupta (2002) argue that "[a]lthough not necessarily immune to corruption, electoral democracies have been found to foster a vigilant civil society, increased government accountability, and a higher degree of transparency."

The meta-narrative in this strand of the literature is that governments can be non-benevolent and rent-seeking and, therefore, institutions have to exist to make them more accountable to the citizens (Aidt et al. 2008). In particular, the institutions should prevent expropriation of private citizens by a non-benevolent state. In this context, it is easy to see how well-defined property rights, rule of law and democracy (which also guarantees voice to the citizens) can restrict a non-benevolent government's ability to seek rent and expropriate private citizens. How would these institutions emerge, however? One cannot depend on rising income levels alone to facilitate creation of appropriate institutions through greater citizens' demand for good governance. As argued by Kaufmann and Kraay (2002), there are a "variety of reasons why entrenched elites in a small country benefit from the status quo of misgovernance and can successfully resist demands for change even as incomes rise over very long periods of time" (p. 4).[6] For example, aid-driven reforms that aimed to redress the policy bias in favor of urban areas of Africa were resisted and blocked by urban interest groups, among other places, "in Zambia, where riots hastened the reversal of fiscal and exchange rate reforms in the late 1980s" (Ndulu and O'Connell 1999: p. 56).

This, in turn, raises the question as to whether governance quality can be influenced by external intervention, especially for contexts in which domestic political economy makes results in an equilibrium that is characterized by weak governance or misgovernance. Limited evidence from the literature suggests that foreign influence, in the form of foreign direct investment, can help improve domestic institutional environment, for example, reduce corruption (Kwok and Tadesse 2006) and encourage stronger property rights (Ali et al. 2011). However, while it is obvious as to why foreign private economic agents such as multinational enterprises would lobby for (more generally, encourage) stronger property rights or less corruption, it is unclear as to why they would lobby for other aspects of good governance such as greater

[5]Huntington (1968) has argued that "[t]he differences between democracy and dictatorship are less than the differences between those countries whose politics embodies consensus, community, legitimacy, organization, effectiveness, stability, and those countries whose politics is deficient these qualities" (p. 1). Hence, the political structure of a country should not matter per se but, as we highlight shortly, these conditions are more (though perhaps not exclusively) likely to manifest themselves in a democracy in which governments (and the associated political elite) are accountable.

[6]The political economy processes through which political and economic elite create and protect rent are much discussed in the literature. For a discussion of the political economy of financial development, for example, which is often resisted by the political and financial elite, see Bhaumik (2018).

inclusivity and a stronger democratic ethos and voice for the country's citizens.[7] For foreign intervention to work, therefore, the push for better governance may have to come from organizations such as state and non-state donors of foreign aid.

Brautingam and Knack (2004) examine the impact of aid on governance in the context of Sub-Saharan Africa. They acknowledge that aid can have two very different impacts on governance quality. On the one hand, aid can improve governance by "improving the quality of the civil service, strengthening the quality of the civil service, and establish[ing] strong central institutions" (p. 260). On the other hand, multiplicity of aid-funded projects can create conditions such as competitive poaching of able civil servants by projects that are inimical to good governance. Prospect of good governance can be further deterred if aid provides aid-recipient countries with a budgetary slack such that it can avoid greater scrutiny by their citizens,[8] and if politicians and bureaucrats alike treat aid as a fungible common pot of money that they can use to finance their pet projects at no cost to themselves. They find that, in the Sub-Saharan context, high levels of aid were associated with a decline in governance quality. Rajan and Subaramanian (2007) test the proposition that aid might actually facilitate weaker governance, using a sample of all developing countries included in the UNIDO database. They find that "in a country that receives more aid, governance dependent industries grow relatively slower" (p. 325). They conclude that aid may be associated with weak governance because the budgetary slack of the recipient governments enables them to disengage from the citizens—aid reduces the need to tax the citizens—and correspondingly the government is less accountable.

This leaves on the table external shocks such as changes in a country's terms of trade or global competitiveness because of changes in relative prices of tradeable goods in the global market or because of reduction in trade barriers as a consequence of global treaties. Following Rodrik (1999), Ndulu and O'Connell (1999) argue the following:

> Efficient adjustment requires that resources be reallocated to minimize the decline in permanent income, and that total spending decline accordingly. Rodrik argues that the willingness of domestic groups to accommodate such efficiency-based responses depends on beliefs about how conflicting claims will work themselves out. If the political system is believed to reward pressing aggressive claims on resources, then everyone will make such claims, and a tangle of mutually incompatible claims will paralyze effective policy action and exacerbate losses. What matters in the end, in this view, is the strength and independence of policy institutions relative to latent social conflicts. (p. 60).

It is easy to see how, in such circumstances, inclusivity, voice and collective action can be undermined in equilibrium. Once again, therefore, an external shock

[7] At the risk of over-emphasis, it is possible to argue that inclusion, non-discrimination and voice are highly correlated with the much discussed concept of *capability* that has implications for social justice. See Robeyns (2005) for a survey of the relevant literature.

[8] The implications of a soft budget constraint for efficient allocation of resources, collective action and quality of effort are well understood in the wider economic literature (e.g., Kornai 1986; Qian and Roland 1998; Qian and Xu 1998). But much of this discussion is in about enterprises, especially state-owned enterprises, and relationships between federal and regional or local governments, rather than about states and federal governments themselves.

is unlikely to be a panacea that facilitates good governance; domestic sociopolitical dynamics matter.

A recurrent theme in this literature is that weak governance may exist in equilibrium simply because it protects the rents of powerful political and economic interest groups. In come contexts, it may be in the interest of these groups to improve those aspects of governance that reduce the transactions cost of productive activities and the risk of *ex post* expropriation of private producers by the state. However, precisely to preserve their rents, generally by way of control over resources, the political and economic elite in these contexts may be less willing to improve other aspects of governance that make economic growth inclusive and/or give voice to equally all citizens. Indeed, even economic governance may be at peril if the rate of time discount of these interest groups is high, such that rapid expropriation is their best strategy. A discussion of governance, therefore, requires a closer examination of the relationship between the state and private citizens (more broadly, private agents), as well as an idea about the nature of the interest groups that compete for influence in any given context. This is the basis for the framework outlined in the next section.

3 Framework for Examining Governance

Developing a framework to examine governance requires an understanding the role that governance plays in the lives of private citizens. The economic approach to governance emphasizes the role of good governance in facilitating transactions which lies at the heart of specialization and exchange that characterizes modern economies. It is well understood that well-established rules of the game—institutions, formal or informal—are necessary to facilitate transactions, but rules alone are not sufficient; one or both parties can cheat if the rules are not enforced properly. While private citizens (or economic agents) can approach the legal system to adjudicate in the event of disputes, the legal system would require sufficient information to proceed with the process of adjudication and they would be bound by the legislations and decrees that determine the punishment structures. Finally, they would need credibility to ensure that the disputing parties accept their judgment about how to settle the dispute. In this setup, governance can be viewed as the process of information intermediation or enforcement intermediation.

Dixit (2003) examines the implications of these two forms of intermediation in a setup of asymmetric two-sided prisoners' dilemma where two parties transact with each other and, given the information asymmetry associated with most (if not all) economic transactions, both parties have opportunities to cheat. Each party is composed of individuals who are either *Honest* and who play honest unless they have to cheat to punish a deviant player, or *Cheats* who always cheat, or *Opportunists* whose actions depend on the net costs/benefits of cheating. Dixit demonstrates that, in this setup, "[p]rivate intermediation may be infeasible just when it is most valuable, because the prospect of a large profit creates a fierce contest to acquire monopoly over intermediation, which shortens the horizon of the intermediary and threatens

the conditions needed to sustain good behavior in repeated interactions" (p. 472). More importantly, Dixit's framework raises some bigger questions about the nature of governance:

> The members of any large group are engaged in many different kinds of economic interactions, with different payoffs, different possibilities of cheating, and different forms of repetition …. Will a common system of governance cope with them all, or will different modes have different comparative or absolute advantages for governing different transactions? Can there be multiple equilibria, where either mode can persist once it gets established, even though another may have better properties? Should one expect a shift from one mode to another during the process of economic growth or expansion of trade ….? Can the existence of multiple equilibria give rise to lock-in? (p. 472)

Dixit's (2003) framework allows for expropriation of the parties engaged in transactions by the intermediary, which may be a private entity such as the mafia but which, for our purposes, is the government or the state. This goes beyond the paradigm in which rational political elites passively adjust the nature and enforcement of laws and regulations in response to lobby by private interest groups. In the words of Levi (1988), "actors who compose the state have interests of their own, derived from and supported by institutional power. Rulers may sometimes, even often, act on behalf of others. Nonetheless, they are not simply handmaidens of the dominant economic class or other influential actors. They will act in their own interests when they can." (pp. ….). In Levi's framework, the objective of these rulers or the political elite is to maximize revenue—a significant proportion of which is arguably political rent that is used for the private benefits of the rulers—and they enter into different contracts with different groups of private citizens to share the resources that help generate revenue. Understandably, the share of any citizen or interest group is determined by its bargaining power vis-à-vis the rulers, and the bargaining power of the rulers increases with the concentration of coercive, economic and political resources in their hands. It follows that rulers have an incentive to set up governance structures that maximize their own bargaining power. It is now easy to see why governance may be viewed as the mechanism that constrains the power of the political elite or non-benevolent governments to expropriate private citizens (Aidt et al. 2008).[9]

It is also not difficult to understand why private citizens may form groups that compete with each other. As suggested by Fukuyama (2011), "the primordial form of human social organization was tribal" (p. 52), and human societies were probably organized along the lines of bands—a predecessor of tribes—as far back as hunter-gatherer times. The basis for the formation of these tribes or units (which we may label "interest groups") was generally common descent, usually through male lineage,

[9]Note that the possibility of expropriation of private citizens by predatory political regime—more generally, the "state"—does not imply that, as argued by some libertarians, that the size of the state should be arbitrarily small. In order to govern well, a state and its machinery should have "the ability to make and enforce rules, and to deliver services, irrespective of whether that government is democratic or not" (Fukuyama 2013: pp. 3). However, the likelihood and extent of expropriation can simultaneously be limited by measures such as separation of powers between the executive, the legislature and the judiciary, judicial independence, and strict term limits for presidents and other members of the executive branch.

and kinship may have been cemented by the emergence of religions that bound the individuals with each tribe by common beliefs. These interest groups were often engaged in violent conflicts and in these conflicts lay the genesis for the formation of states (North et al. 2009; Fukuyama 2011). In particular, coalitions of tribes could emerge to guarantee security to private citizens, in return for monopoly over coercive powers. Once formed, a state could also act as a third-party arbitrator in the event of conflicts between the constituent tribes (and, hence, interest groups), and the monopoly over coercive powers made its claims to enforcing rules and laws using the state machinery credible.

However, since the political elites alone cannot prevent violence both within and between states, they have to co-opt groups of citizens through distribution of resources and rent-sharing agreements. In North et al.'s (2009) words: "[n]atural state elites sit at the top of, but are also embedded in, patron-client networks that extend down to the rest of the society. …. [p]atron client relationships not only structure the creation, gathering, and distribution of rents that can limit violence; the structures also organize violence itself" (pp. ….). With evolution, the natural state is able to initially support public institutions connected to the state that structure the shared beliefs of the political elite within the dominant coalition emerge, and eventually institutions that support elite organizations outside the immediate domain of the state.[10] The centrality of elite privileges in a natural state, irrespective of its stage of evolution,[11] is easy to comprehend from this brief discussion about the characterization of these states, with attendant implications for governance quality in these states.

The characterizations of governance discussed earlier, whether in terms of limits to the state's (or the government's) ability to expropriate private citizens, or in terms of specific attributes such as democracy and voice, coincide with what North et al. (2009) call an *open access order*.

> All open access societies satisfy the Weberian assumption: their states possess a monopoly on the legitimate use of violence. Consolidation of violence carries the danger of the state using violence for its own ends. As a result, the logic of controlling violence in the open access order involves three elements: 1) consolidated organization of military and policy forces is subject to the control of the political system; 2) the political system must be constrained by a set of institutions and institutions that limit the illegitimate use of violence; and 3) for a political faction or party to remain in power, it must enjoy the support of economic and social interests, broadly defined. …. Control of the political system is open to entry by any group and contested through prescribed, and typically, formal constitutional means. All citizens have the right to form organizations, and they use the services of the state to structure the internal and external relationships of these organizations to individuals and other organizations. (pp. ….)

In open access orders too individuals and organization pursue rent, for example, through corporate strategies that erects barriers to market entry, but rents are eroded rapidly on account of political and economic competition. Nor are political and

[10] For a discussion about the role and importance of political institutions, see Huntington (1968).

[11] In North et al.'s (2009) taxonomy, these stages of evolution are *fragile* natural state, *basic* natural state and *mature* natural state.

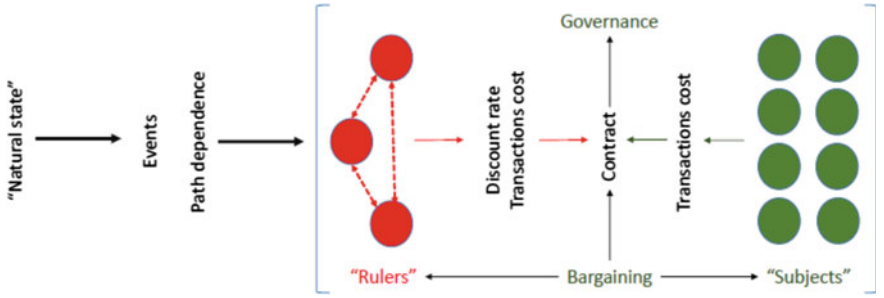


Fig. 3 Framework for examining governance

economic systems independent of each other in these orders; for example, an open economic order is necessary to sustain an open political order. More importantly, each individual citizen has an equal right to pursue rents and to compete with incumbent rent-seekers, and the state and the political elite associated with it are as accountable as private citizens. It is evident, from measures of governance quality (however noisy) and related measures of state attributes such as democracy that only a handful of countries can be labeled open access orders, and this transition is not automatic by any stretch of imagination.

We now have the elements of a framework that can be used to examine governance and, in keeping with the sentiment expressed in Dixit (2009), any discussion of governance may have to be context-specific in order to be meaningful, but the discussion in each context may be feasible using a common framework. As highlighted in Fig. 3, any discussion about governance quality in a given context may have to start from a discussion of its historical "natural state," to identify the interest groups, the constituents of the dominant coalitions and the sociopolitical compact both among the elites and between elites and the other citizens. The identities of the contemporary elites and interest groups may be the outcome of path dependency or may be radically altered by events such as domestic-civil or cross-border wars, major changes to the political and economic systems such as those observed in Central and Eastern Europe and the former Soviet Republics during the 1990s, and major changes driven by legislation such as the empowerment of the *other backward classes* in India since the late 1980s.

The understanding of this evolutionary process is particularly important to establish important parameters such as trends in the relative bargaining power of the contemporary rulers (i.e., political elite) and subjects (i.e., private citizens), the rate of time preference of the political elite, and whether there is sufficient trust among the different interest groups to facilitate meaningful bargaining (i.e., the transactions cost associated with any contract). Given these parameters, one can have a meaningful discussion about a social contract, the equilibrium nature of which will be characterized by the aforementioned parameters. A discussion of the possible ways to formalize the bargaining process lies outside the scope of this paper; see Binmore (1994, 1998).

4 Concluding Comments

Data suggest that for most countries good governance is an aspiration, with little guarantee that it would become reality in the foreseeable future. In the same vein, North et al. (2007) suggest that only have a handful of countries made the transition to open access orders. Further, evidence discussed earlier in this chapter suggests that economic growth is possible even when a country does not make significant progress toward good governance, as reflected in measures of rule of law and other components of governance. Indeed, the discussion in the literature suggests that it may be reasonable to shift our focus from all-encompassing progress in governance quality to the minimum governance level that is necessary for political and economic progress (Grindle 2007). Why then should we be interested in governance?

The answer lies in North et al. 's (2009) observation that, for a country, there is also no guarantee of progression along the spectrum that ranges from a fragile natural state to an open access order—states can move both forward and backward in terms of political (and economic) development. While it is easy to create formal institutions that are consistent with open access orders and good governance—independent judiciary, separation of powers between the executive and other branches of the state that can make the executive accountable, and rights of citizens enshrined in law, to name a few—there are significant informal barriers that limit access to these institutions for all but the political (and economic) elite (North et al. 2007). Evidence from contexts such as Africa suggests that incumbent rent-seekers, who may either be elites themselves or may be in coalition with elites, do indeed oppose changes that can facilitate better governance when better governance leads to dissipation of rent and private benefits.

More worryingly, the recent experience of a wide range of countries around the world suggests that countries can indeed move backward in key metrics associated with good governance. For example, on the measure of democracy reported by the Economist Intelligence Unit—a 0–10 scale on which 8–10 corresponds to full democracy and 0–4 corresponds to authoritarian regimes[12]—Hungary's score declined from 7.53 in 2006 to 6.72 in 2016. The corresponding figures for some other countries are as follows: 7.30 and 6.83 for Poland, 7.91 and 7.41 for South Africa, 5.70 and 5.04 for Turkey, 8.22 and 8.98 for the USA, and 5.42 and 4.68 for Venezuela.[13] At the same time, factors such as economic inequality is paving the way for populism in countries such as the USA and Europe (Inglehart and Norris 2016; Algan et al. 2017); by their very nature, this variety of populism is inimical to inclusiveness.

A better understanding of the dynamics that shapes governance is, therefore, necessary not only to construct institutions that can withstand interest group pressures and temporary disruptions to the political and economic systems, but also to better understand whether or under what conditions a country might move backward on the governance scales. At present, the latter is possibly no more than speculation in

[12]Source: https://infographics.economist.com/2017/DemocracyIndex/.

[13]For some of these countries, detailed discussions of institutional reversals, with attendant implications for governance quality, are available. See, for example, Acemoglu and Ucer (2015).

the developed open access order world (e.g., Mickey et al. 2017), but tougher days for good governance (and open access orders) and all that are associated with it may lie ahead.

Acknowledgements The author would like to thank Jeffrey Nugent, Timur Kuran, Pauline Dibben, and workshop participants at the 2015 summer school of the Leibniz Institute for East and Southeast European Studies (IoS), Regensburg, Germany, and the 2017 Development Economics Conference at the University of Lincoln, UK, for useful feedback on early thoughts on governance. He remains responsible for all remaining errors.

References

Abed, G. T., & Gupta, S. (2002). The economics of corruption: An overview. In G. T. Abed & S. Gupta, (Eds.), *Governance, corruption, & economic performance*. Washington, DC: The International Monetary Fund. http://www.imf.org/external/pubs/nft/2002/govern/#chp1.

Acemoglu, D., & Ucer, M. (2015). *The ups and downs of Turkish growth, 2002–2015: Political dynamics, the European Union and the institutional slide*. Working paper no. 21608, National Bureau of Economic Research, Cambridge, MA.

Aidt, T., Dutta, J., & Sena, V. (2008). Governance regimes, corruption and growth: Theory and evidence. *Journal of Comparative Economics, 36*(2), 195–220.

Algan, Y., Guriev, S., Papaiannou, E., & Passari, E. (2017). The European trust crisis and the rise of populism, BPEA conference drafts. *Brookings Papers on Economic Activity*, Downloadable from https://www.brookings.edu/wp-content/uploads/2017/09/4_alganetal.pdf.

Ali, F., Fliess, N., & Macdonald, R. (2011). Climbing to the top? Foreign direct investment and property rights. *Economic Inquiry, 49*(1), 289–302.

Berkowitz, D., Pistor, K., & Richard, J.-F. (2003). Economic development, legality, and the transplant effect. *European Economic Review, 47*(1), 165–195.

Bhaumik, S. K. (2018). The political economy of financial development: A review. In Y. Dwivedi et al. (Ed.), *Emerging markets from a multidisciplinary perspective—Challenges, opportunities and research agenda*. Springer.

Binmore, K. (1994). *Game theory and the social contract, playing fair* (Vol. 1). Cambridge, MA: The MIT Press.

Binmore, K. (1998). *Game theory and the social contract, Just Playing* (Vol. 2). Cambridge, MA: The MIT Press.

Brautingam, D. A., & Knack, S. (2004). Foreign aid, institutions, and governance in Sub-Saharan Africa. *Economic Development and Cultural Change, 52*(2), 255–285.

Dixit, A. (2003). On modes of economic governance. *Econometrica, 71*(2), 449–481.

Dixit, A. (2009). Governance institutions and economic activity. *American Economic Review, 99*(1), 3–24.

Fukuyama, F. (2011). *The origins of political order: From prehuman times to the French revolution*. London: Profile Books Limited.

Fukuyama, F. (2013). What is governance? *Governance, 26*(3), 347–368.

Grindle, M. S. (2007). Good enough governance revisited. *Development Policy Review, 25*(5), 553–574.

Huntington, S. P. (1968). *Political order in changing societies*. New Haven, CT: Yale University Press.

Inglehart, R. F., & Norris, P. (2016). *Trump, Brexit, and the rise of populism: Economic have-nots and cultural backlash*. Faculty research working paper no. RWP16-026, John F. Kennedy School of Government, Harvard University.

Kaufmann, D., & Kraay, A. (2002). *Growth without governance.* Policy research working paper no. 2928, The World Bank, Washington, DC.

Kaufmann, D., Kraay, A., & Mastruzzi, M. (2010). *The worldwide governance indicators: Methodology and analytical issues*. Policy research working paper no. 5430, The World Bank, Washington, DC.

Knack, S., & Keefer, P. (1995). Institutions and economic performance: Cross-country tests using alternative institutional measures. *Economics and Politics, 7*(3), 207–227.

Kornai, J. (1986). The soft budget constraint. *Kyklos, 39*(1), 3–30.

Kurtz, M. J., & Schrank, A. (2007). Growth and governance: Models, measures and mechanism. *Journal of Politics, 69,* 538–554.

Kwok, C. C. Y., & Tadesse, S. (2006). The MNC as an agent of change for host-country institutions: FDI and corruption. *Journal of International Business Studies, 37*(6), 767–785.

Levi, M. (1988). *Of rules and revenue*. Berkeley, CA: University of California Press.

Mickey, R., Levitsky, S., & Way, L. A. (2017). Is America still safe for democracy? Why the United States is in danger of backsliding. *Foreign Affairs* 20–29.

Ndulu, B. J., & O'Connell, S. A. (1999). Governance and growth in Sub-Saharan Africa. *Journal of Economic Perspectives, 13*(3), 41–66.

North, D. (1987). Institutions, transactions costs and economic growth. *Economic Inquiry, 25*(3), 419–428.

North, D. (1989). Institutions and economic growth: An historical introduction. *World Development, 17*(9), 1319–1332.

North, D. C., Wallis, J. J., Webb, S. B., & Weingast, B. R. (2007). *Limited access orders in the developing world: A new approach to the problems of development*. Policy research working paper no. 4359, The World Bank, Washington, DC.

North, D. C., Wallis, J. J., & Weingast, B. R. (2009). *Violence and social orders: A conceptual framework for interpreting recorded human history*. New York, NY: Cambridge University Press.

Qian, Y., & Roland, G. (1998). Federalism and the soft budget constraint. *American Economic Review, 88*(5), 1143–1162.

Qian, Y., & Xu, C. (1998). Innovation and bureaucracy under soft and hard budget constraints. *Review of Economic Studies, 65*(1), 151–164.

Rajan, R., & Subramanian, A. (2007). Does aid affect governance? *American Economic Review, 97*(2), 322–327.

Rivera-Batiz, F. (2002). Democracy, governance, and economic growth: Theory and evidence. *Review of Development Economics, 6*(2), 225–247.

Robeyns, I. (2005). The capability approach: A theoretical survey. *Journal of Human Development, 6*(1), 93–114.

Rodrik, D. (1999). Where did all the growth go? External shocks, social conflict, and growth collapses. *Journal of Economic Growth, 4*(4), 385–412.

Rodrik, D., Subramanian, A., & Trebbi, F. (2004). Institutions rule: The primacy of institutions over geography and integration in economic development. *Journal of Economic Growth, 9*(2), 131–165.

From Political Economy to Economics and Back Again?

Amitava Krishna Dutt

1 Introduction

Once upon a time there was a field of study called "political economy". Many books on what has been called the economy were written in the field, and a fair number had the term in their title. However, by the end of the nineteenth century the term started to go out of fashion, and the study of the economy increasingly came to be called economics, which is still the overwhelmingly popular term. However, the use of the term political economy has increased in the last several decades, although it means different things to different people. The purpose of this paper is to review the history of this shift, what was and is meant by the terms political economy and economics, and the reasons for the change in the name and nature of the field, with a view to examining whether a return to the term political economy instead of economics is desirable. It will be argued that a return to the name political economy from economics and a return to what political economy tried to do before the change in name is desirable not only for a better understanding of the economy, but also for the well-being of people, especially those who have been excluded and marginalized.

The rest of this paper proceeds as follows. Section 2 describes the change in the name from political economy to economics. Section 3 discusses the revival of the use of the name political economy is the last few decades. Section 4 examines different views on what political economy means and what distinguishes the term

Since Anup Sinha (or Anupda) was, for me, an early influence pushing me to value both the rigorous analysis of the economy and the importance of political economy, I am very pleased to contribute this paper in a book put together in his honor.

A. K. Dutt (✉)
Department of Political Science, University of Notre Dame, Notre Dame, USA
e-mail: adutt@nd.edu

A. K. Dutt
FLACSO, Quito, Ecuador

P. Ray et al. (eds.), *Economics, Management and Sustainability*,
https://doi.org/10.1007/978-981-13-1894-8_3

from economics. Section 5 provides a brief history of thinking on the economy with a view to providing a broader view on how the field changed *pari passu* with the change in its name. Section 6 analyzes some problems with what economics has become by distinguishing itself from political economy, and Sect. 7 explores possible reasons for the change in both name and content of the discipline. Section 8 ties the argument of the different sections of the paper to make the case for a return to political economy.

Before proceeding further we should clarify the meaning of three terms that will be used in this paper (one of which has already been used), that is, the economy, the polity and society. All three will be used in two related senses, one referring to a system and the other to some aspects of human actions or behavior. First, in the sense of systems, following Samuelson's introductory text, the economy can be defined as a system which answers the following questions about the production of goods and services: what, how and for whom, that is, what is produced, how it is produced, and who gets the goods and services produced.[1] In the sense of behavior, the economy can be considered to be a site in which people take actions that are "most closely connected with the attainment and with the use of the material requisites of well-being" (Marshall 1890), involving production, consumption, work, purchase and sale, saving, investing (that is, installing new capital goods), and holding assets such as financial assets. Second, the polity is a system by which a society as a whole is governed and which guides the processes by which the government is selected and affected by people and groups, and the site in which activities of the state and its components is conducted, and in which the activities of people and groups in selecting and influencing the state are conducted (such as governing, providing collective goods and services, interacting with other states internationally, voting, rebelling, and working for political parties and campaigns). Third, society is a system in the context of which people and groups identify themselves, act, and interact with each other. A broad definition of society includes the economy and the polity, but in a narrow sense it can be taken to refer to those elements of the system and those activities that are not typically included in the economy and the polity; we will use the term to refer to both senses. These definitions are often identified with the disciplines of economics, politics and sociology and, indeed, the definitions have been compiled from some standard definitions of these subjects, such as "economics is the study of the economy or economic behavior." However, as we shall see, we will make no such identification and, to the contrary, have reason to question it, especially in Sect. 6.

2 Use of the Terms Political Economy and Economics

The term political economy is derived from the Greek words οἰκονομία (oikonomia) or household management and *polis* (polis) or city-state, and literally means city-

[1] See, for instance, pg. 10 of the 1998 16th edition of Samuelson (1948), coauthored by Paul Samuelson and William Nordhaus.

state management.[2] The term seems to have first appeared in the title of a book by Montchretien (1615), although it was used earlier by Louis de Mayerne-Turquet in 1611 in the context of discussing the duties, including those involving what can be called economic matters, of the sovereign power toward the citizens of the state it ruled (see King 1948). In Italy, Verri (1763) used the term *Economia politica* to refer to the production and distribution of wealth in the context of the management of a nation's resources. In the English-speaking world the term was first used in the title of Sir James Steuart's *Principles of Political Oeconomy*, which states "What oeconomy is in a family political oeconomy is in a state" (Steuart 1767, 2). Though Smith (1776) did not use it in the title of *The Wealth of Nations*, he made it clear that what he was addressing was the subject of political economy, which he describes in Book IV as "a branch of the science of a statesman or legislator .. [which has] … two distinct objects; first, to provide a plentiful revenue or subsistence for the people, or, more properly, to enable them to provide such a revenue or subsistence for themselves; and, secondly, to supply the state or commonwealth with a revenue sufficient for the public services. It proposes to enrich both the people and the sovereign" (Smith 1776: 678–9). After Smith, the major British classical economists, including Ricardo (1817), Malthus (1820), and Mill (1844) all used the term political economy in the titles of their major works on the economy.

The shift to the term economics was first prominently made by Alfred Marshall, initially in his book *Economics of Industry* coauthored with his wife (Marshall and Marshall 1879) and then in his *Principles of Economics* (Marshall 1890) although the term appeared slightly earlier in books by Sturtevant (1877) and MacLeod (1878). Marshall and Marshall (1879: 2) explained that they preferred drop the term "political economy" because "political interests generally mean the interest of some part or parts of the nation," referring to political narrowly to factional interests and excluding the interests of a nation. However, when it comes to defining economics in the *Principles*, Marshall (1890, 1) does not distinguish between political economy and economics, writing that "Political Economy or Economics is a study of mankind in the ordinary business of life…". Groenwegen (1985, 748) argues that the Marshalls replaced political economy by economics in the first book because they did not want incorrected inferences to be drawn regarding "the labour question" and the movement of the working classes, since he rejected what he thought were radical political solutions, supporting, rather the extension of educational opportunities. The name also fit in with Marshalls's focus on the scientific nature of the subject. What is surprising therefore, is Marshall (1890) treating the two terms synonymously, and Groenwegen (1985) explains it in terms of his desire to enhance the study of economics at Cambridge by emphasizing its long pedigree as political economy.

Although Menger (1871) was translated in the 1950s as *Principles of Economics*, the name he used for the science was *Volkswirtschaftslehre* which is literally translated as the "study of national household management." Jevons's (1871) *Theory of Political Economy* retained the title in later editions, but in the second edition he explained, regarding "the substitution for the name Political Economy of the single

[2]The following discussion draws heavily on Arndt (1984a, b) and Groenwegen (1985, 1987).

convenient term *Economics*" that "it would be well to discard, as quickly as possible, the old troublesome double-worded name of our Science. Several authors have tried to introduce totally new names, such as Plutology, Chrematistics, Catallactics. But why do we need anything better than Economics? This term, besides being more familiar and closely related to the old term, is perfectly analogous in form to *Mathematics*, *Ethics*, *Æsthetics*, and the names of various other branches of knowledge, and it has moreover the [xiv] authority of usage from the time of Aristotle. Mr. Macleod is, so far as I know, the re-introducer of the name in recent years, but it appears to have been adopted also by Mr. Alfred Marshall at Cambridge. It is thus to be hoped that *Economics* will become the recognised name of a science … Though employing the new name in the text, it was obviously undesirable to alter the title-page of the book." (Jevons 1879, xiii–xiv). Leon Walras's 1874 original French version of his classis work was entitled *Éléments d'économie politique pure*, but Jaffe's English translation made it *Elements of Pure Economics* (see Walras, 1954).

Knut Wicksell (1901–6), stated that while "political economy" was appropriate for the mercantilists, for whom the state's duty was seen a influencing the affairs of individuals, it became less so with the advent of Physiocracy and the rise of the doctrine of *laissez-faire*. He, nevertheless retained the old term by naming his book *Lectures in Political Economy*, because "in accordance with the modern outlook, the subject matter of political economy is becoming more and more the doctrine of economic phenomena in their interrelations, seen *as a whole*" (p. 1), and "As soon as we begin seriously to regard economic phenomena *as a whole* and to seek for the conditions of the welfare of the whole, consideration of the interests of the proletariat must emerge; and from thence to the proclamation of equal rights for all is only a short step. The very concept of political economy, therefore, or the existence of the science with such a name, implies, strictly speaking, a thoroughly revolutionary programme" (p. 41). Groenwegen (1985) points out that Wicksell was hardly an exception in retaining the old name, mentioning Sidgwick (1883) with subsequent editions in 1887 and 1901, Wicksteed who has been referred to as the purist of the marginalist school, published *The Common Sense of Political Economy*, and Edgeworth (1925), who was closely associated with Marshall, published his collected essays with the name *Papers Relating to Political Economy*. Even Pareto, the most famous Italian Walrasian, published his 1897–98 text under the title *Cours d'economie politique* and his second text in 1907 as *Manuale d'economia politica.* All these works are devoted to pure economic theory in a narrow sense.

Eventually, however, the term political economy was replaced by the term economics following the lead of Marshall and others, and the hopes of Jevons. In fact, as we shall see in Sect. 4, economics came to be distinguished from something *different*, called political economy. Schumpeter (1954, p. 36–43) distinguishes "economic analysis" interpreted as value-free science (and in which it is possible to gauge whether or not there has been scientific progress) from "systems of political economy" interpreted as "a comprehensive set of policies" based on "certain unifying (normative) principles such as the principles of economic liberalism, of socialism, and so on." However, the term subsequently fell into disfavor and its used became rare. In the five years from 1953 to 1957, only three books were published in the

English-speaking world containing the term in its title (Whynes 1984: 1). Although a highly ranked and influential mainstream economics journal, *Journal of Political Economy*, which was founded in 1892, still carries the term in its name,[3] mainstream (and even some heterodox) economics journals all have the term economics in their titles and introductory texts carry the same title. Even introductory texts that explicitly try to provide heterodox perspectives, such as Fusfeld (1972) and Robinson and Eatwell (1973) use the title economics.

3 The Revival of Political Economy

From the late 1960s, however, the term political economy experienced a revival. Staniland (1985: 1) writes that "a quick check of a university library catalog recently revealed some 117 books so titled, covering subjects as diverse as advertising, Appalachia, art, drug trafficking, East-West trade, human rights, independent Fiji, Turkish income distribution, Nasserism, the space program, indirect rule in Mysore, slavery, war, racism and Pondoland." A number of journals appeared with the old name, including (with their founding year) *Review of Radical Political Economics* (1969), *History of Political Economy* (1969), *Review of Black Political Economy* (1970), *International Journal of Political Economy* (1971), *Review of African Political Economy* (1974), *Studies in Political Economy* (1979), *Contributions to Political Economy* (1982), *European Journal of Political Economy* (1985), *Review of Political Economy* (1989), *Review of International Political Economy* (1994) and *New Political Economy* (1996). A Google Scholar search of political economy yielded 3,030,000 results, as compared with 4,360,000 for economics, 3,850,000 for sociology, 2,740,000 for political science and 2,850,000 for politics. Even a number of textbooks and discussions of the subject, began to appear, such as Phelps (1985), although it did not go beyond a first edition unlike many introductory textbooks, Eaton (1966), Caporaso and Levine (1992) and Stillwell (2002).

After the revival of the term political economy, however, it did not mean the same thing to everyone. Two specific and well-known traditions of political economy are the Marxian and neoclassical ones, and another one refers to international or global issues.

Marxian political economy draws on Marx's analysis of distribution, capital accumulation, technical change and crises, but also the role of social classes and the state and political struggle. In broad terms, dialectical materialism allows for a general examination of the reciprocal relation between the economic and social base and the political and ideological superstructure, which serves as a fruitful way of incorporating the economy, polity, and society, into a theory of social change. Narrow versions

[3]The Web site of the journal, however, provides no justification for the name and seems to describe it as an economics journal, stating "One of the oldest and most prestigious journals in economics, the *Journal of Political Economy* (*JPE*) has since 1892 presented significant research and scholarship in economic theory and practice".

of Marxian political economy have sought to take what may be called a more fundamentalist view, in which exploitation and income distribution is seen to arise in the production process, and not in circulation, and in which the base determines the superstructure and not the other way around. Baran (1957) examined the problems of growth and development from a Marxist perspective in his *Political Economy of Growth*, adopting the political economy name because his study required an "understanding of the factors responsible for the size and utilization of the social surplus … a problem not even approached in the realm of pure economics" (Baran 1957, 131).

Neoclassical political economy has sought to apply the economic theory of rational choice, in which individual economic agents maximized their utility in a self-interested manner, to the field of politics, among others. Various strands of this approach exist, their common feature being to use a particular approach to examining the world using the optimizing (or rational), self-interested agent, originally used in economics, to examine social issues, including the marriage, child-rearing, crime, and political issues, including policy formulation and implementation, the writing of laws and constitutions and legislative behavior. Different closely—related strands to this literature have been distinguished, including that of the Chicago school, following Gary Becker, which seeks to apply the rational choice approach to the study of non-market phenomena, the public choice school, following Downs and Buchanan, which focuses on political behavior and more eclectic neoclassical approaches involving, for instance, politometrics, that is, formal models and their statistical testing of political behavior (see Whynes 1984). There are also some similarities between these approaches and the Austrian approach which follows Hayek in focusing on the informational aspects of decentralized exchange and the neo-institutional approach pioneered by North (1991), which focuses, among other things, on transactions costs and how they are reduced by institutions in order to increase economic efficiency.

The term political economy has also been used widely in the study of international relations, in what has been called international political economy or global political economy. As recounted by Cohen (2008), this field combined the analysis of the state and markets, and the study of power and plenty, combining economic and political factors into international relations. The approach, however, is not unified, Cohen himself distinguishing between American and British international political economy, representing different approaches, such as neoclassical and heterodox, and different methods, formal and discursive. Further distinctions can be drawn between those who adopt a realist, liberal-institutionalist, and Marxian-radical approaches (including dependency and world-system ones) to international relations. The journal *Review of International Political Economy* reflects these different approaches much more than journals publishing on international political economy in the USA, such as *International Organization*, which is more aligned to Cohen's American school.

Some writers, however, do interpret political economy in a broad sense, incorporating economics, politics and ethics and/or covering a broad range of economic approaches. Thus, for instance, Caporaso and Levine (1992) is described by the publisher as exploring "some of the more important frameworks for understanding the relation between politics and economics, including the classical, Marxian, Keynesian,

neoclassical, state-centered, power-centered, and justice-centered." Stilwell (2002) is described as making "the complexities of contesting economic ideas—including classical political economy and Marxist economics and neoclassical economics and neo-liberalism—clear and accessible to students." The *European Journal of Political Economy* and *New Political Economy* are also eclectic, although the former leans more toward neoclassical political economy and the latter to heterodox social science.

4 Economics Versus Political Economy

After the spread of the term "economics," several leading economists attempted to distinguish between the new term and the old term, "political economy." Several other attempts at distinguishing between the two emerged after the revival of political economy in more recent years.

First, an early approach to the difference drew on John Stuart Mill's (1844), distinction between the "science" and the "art" of political economy where the former referred to a "body of truths which had … acquired a collective denomination" which consists of a collection of truths about how a nation becomes rich and the latter to prescriptive and normative issues, a body of practical rules that say "Do this. Avoid that." While Mill took both science and art to be included in political economy, later writers seem to identify the science with "economics" and the art with "political economy." Thus, as noted earlier, Schumpeter (1954, pp. 36–43) distinguishes "economic analysis" interpreted as value-free science from "systems of political economy" interpreted as "a comprehensive set of policies based on certain unifying (normative) principles." Lionel Robbins (1932) also seems to view economics as "as a system of theoretical and positive knowledge" (Fraser 1937, 30), when he defined economics as "the science which studies human behaviour as a relationship between ends and scarce means which have alternative uses" (Robbins 1932, p. 15), and reserves the term political economy for applied topics, including monopoly, protection, planning and government fiscal policy, which are subjects included in his essays on political economy (Robbins 1937). In his later work, Robbins returns to the same distinction, distinguishing between economics as a positive science (though not the same as the natural science, since it deals with rational actors, and although he seems to place welfare economics within it presumably while it does involve making value judgments, it concentrates on their implications) and political economy, which is "concerned with the assumptions of policy and the results following from them" (Robbins 1981, p. 8).

Second, Groenwegen (1987, 906) argues that the two terms, "political economy" and "economics," "have both experienced changes in meaning. Nevertheless, they can still essentially be regarded as synonymous, a feature of this nomenclature reflecting an interesting characteristic of the science it describes. In its sometimes discontinuous development, economics and political economy has invariably experienced difficulties in discarding earlier views, and traces of old doctrine are intermingled

with the latest developments of the science." Among more recent discussions of the distinction between economics and political economy, Arndt (1991, 115–6) sees political economists as being "committed fighters in one good cause or another" and having "at the back of their minds, some vision of a society which they would help to bring about." Groenwegen (1992) disagrees, arguing that Marshall, who had a role in the shift from political economy to economics, had a clear vision of a good society, when for instance, he says that "Taking for granted that a more equal distribution of wealth is to be desired, how far would this justify changes in the institutions of property, or limitations of free enterprise when they would be likely to diminish the aggregate of wealth? In other words, how far should an increase in the income of the poorer classes and a diminution of their work be aimed at, even if it involve some lessening of national material wealth?" (Marshall 1961, 41–2). Clearly, Marshall seems to be raising important normative issues, although it is not clear that raising these questions makes him a committed fighter for a good cause. One may also note that Marshall's analysis was broader in scope in emphasizing evolutionary changes in addition to equilibrium positions, and followed a less rigorous deductive approach than say Walras and Jevons and subsequent neoclassicals. However, Groenwegen seems to overstress the case of Marshall, given that he has worked extensively on Marshall. Moreover, changes in terms in title do not occur at a single point in time, since longer lags due to habit and inertia can be expected.

Third, Robert Gilpin, one of the founders of international political economy, defines political economy by distinguishing between economics and politics and by taking political economy to combine these two fields. Thus, Gilpin (1975: 43) defines political economy as "the reciprocal and dynamic interaction … of the pursuit of wealth and the pursuit of power," focusing on the *objective* of the activities of "economics" and "politics" either for individuals and groups in society, or of the analysis itself. Gilpin here follows a long tradition distinguishing the goals of power (a political one) and wealth (an economic one). Viner's (1948) examination of the mercantilist writers in the seventeenth and eighteenth centuries in Europe suggests that they and the policymakers were driven by the goals of both "power" and "plenty" (of wealth), thereby having both the political goal of power and the economic goal of wealth. In this sense, by Gilpin's definition, the mercantilists were involved in what is the object of study of political economy. In a later work, Gilpin (1987: 8) defines political economy in terms of the state and the market: "[t]he parallel existence and mutual interaction of the 'state' and 'market' in the modern world create 'political economy'; without both state and market there would be no political economy." He goes on to say that "[i]n the absence of the state, the price mechanism and market forces would determine the outcome of economic activities; this would be the pure world of the economist. In the absence of the market, the state or its equivalent would allocate economic resources; this would be the pure world of the political scientist." Gilpin recognizes that there the state and the market as defined here do not exist in pure form and are seen as Weberian ideal types; that economic, political *and social* factors interact; and that the state and the market interact as the embodiment of politics and economics, since the state and its associated political processes affect the production and distribution of wealth, and markets and economic forces affect

the power and welfare of different actors and states. Gilpin is not alone in focusing on the state and the market. Strange's (1988) book on international political economy is entitled *States and markets*.

A fourth current view of political economy distinguishes it from economics in terms of taking a critical or heterodox perspective on the study of the economy. In this view economics is seen as being dominated by neoclassical economics which emerged from the marginalist economics of the 1870s, and with rigorous statistical analysis, that is, with econometric methods. While not necessarily eschewing formalism and econometric methods, political economists are seen as drawing inspiration from classical political economy, which focused on how economic surplus is generated in capitalist economies, how this surplus is distributed, how distribution feeds into accumulation and technical change, and the consequences of this for the economy. It is also related to other heterodox schools of economic thought, including Marxian and radical political economy, institutional economics, post-Keynesian and Kaleckian economics, feminist political economy, among others. These approaches emphasize the importance of power, and as such, incorporate broader political and social phenomena, and therefore often enter into what has been called broader political economy analysis which are analyzed by discursive and ideographic methods, rather than nomothetic ones that attempt to explain all phenomena in terms of law-like tendencies. Some heterodox scholars even go further and seem to reserve the term for their approach. Titles of publications such as *Review of Radical Political Economics* (although, according to its aims and scope, it is open to other heterodox approaches) and *Studies in Political Economy,* which take a Marxian-radical and socialist perspective, and *Contributions to Political Economy* takes a classical-Marxian perspective, adopt a subset of heterodox perspectives, while *Review of Political Economy* and *International Journal of Political Economy* take a more broadly heterodox perspective. However, institutionalist publications, like *Journal of Economic Issues*, the post-Keynesian one called *Journal of Post-Keynesian Economics*, and the general heterodox journal, *Cambridge Journal of Economics*, maintain the term "economics" in their title.

A final distinction between economics and political economy, coming from neoclassical political economy, can be considered to be both a narrowing and a broadening of mainstream economics. It is a broadening in the sense that the economic rational choice framework is extended to deal with political and social issues, as discussed earlier. However, it can also be seen as a narrowing in the sense that it is a branch of mainstream economics that is concerned with political behavior and institutions. This view is reflected by Weingast and Wittman (2006a) in their introduction to the *Oxford Handbook of Political Economy* which belongs to the series called *The Oxford Handbook of Political Science*, who view "political economy' .. [as] the methodology of economics applied to the analysis of political behavior and institutions. As such, it is not a single, unified approach, but a family of approaches … tied together by a set of methodologies, typically associated with economics, but now part and parcel of political science itself. The unit of analysis is typically the individual. The individual is motivated to achieve goals (usually preference maximization but in evolutionary games, maximization of surviving offspring), the theory based in

mathematics (often game theoretic), and the empirics either use sophisticated statistical techniques or involve experiments where money is used as a motivating force in the experiment" (pp. 3–4). They view that this definition to be a "grand (if imperfect) synthesis of … various strands" (p. 3), including the meaning attached to the term by Adam Smith and Karl Marx, and more recent research, an area of study which examines the interrelationship between economics and politics or, alternatively, as a methodological approach that uses economic methodology emphasizing individual rationality or the sociological approach where the level of analysis tends to be institutional rather than the individual.[4]

5 A Very Brief History of Political Economy and Economics

This section provides a very brief overview of the history of thought in political economy and economics. Since the subject matter is very large and nothing but a quick review can be offered in a section of a paper, this section will confine itself to selectively tracing the evolution of thought in this area with a view to examining how the subject narrowed in scope over time (see also Milonakis and Fine 2009 for more details on Western ideas).

We start with early thinking on the issue in Greece and India. For the Greek philosophers politics was the master discipline, the art of arts, which is concerned with the cure of souls or the pursuit of virtue through the proper organization of the polis (Campbell 1987). Economics was not seen as an autonomous inquiry but as being subordinated to the art of arts, interpreted as estate or household management which, according to Aristotle's *Politics* of the fourth century BCE "attends more to men than to the acquisition of inanimate things, and to human excellence more than to the excellence of property which we call wealth" (1259b). Even earlier, Plato emphasized the importance of the division of labor and specialization and of exchange of goods and services for the proper organization of the polis, requiring citizens to do what they do best and only that (Lowry1987). In the fourth and third centuries BCE in India, Kautilya's *Arthashatra*, which has been translated as the "science of politics" by Kangle (see Kautilya 1965) and the "science of political economy" (Boesche 2002), but can be literally translated as "treatise on wealth," provided guidance to the king for ruling a kingdom. This guidance involves political stratagem (for instance, how to deal with and even annihilate political enemies), government administration, and promoting the material well-being of subjects through economic policies (including,

[4]They argue that their definition includes institutions, politics and economics, and hence is a synthesis. They do not say how Smith (who, according to them, defined it as a science of managing a nation's resources to generate wealth) and Marx (who, according to them, viewed it as the study of how ownership of the means of production influences historical processes), fit into this synthesis; perhaps they can be ignored because they are dated. Their definition seems closest to the economics methodology emphasizing individual rationality, narrowed to analyzing political behavior and institutions, and somewhat extended to include evolutionary games and the use of sophisticated statistical techniques (heaven forbid the use of unsophisticated techniques!) and experiments.

for instance, land redistribution). Kautilya argued that the king had to attend to the welfare of subjects as a matter of duty, but also because doing so increases his support among the people and government revenues required for governing the country (see, for instance, Trautmann 2012).

We jump several centuries to the mercantilists, who were active in Europe during the sixteenth to the eighteenth centuries. They emphasized the importance of increasing trade surpluses of states by policies restricting imports and promoting exports. In this, they were concerned both with power and plenty, since a trade surplus brings in precious metals that enable countries to finance their armed forces, especially their navies and increase their military power, and also increase the level of demand for domestic products, thereby increasing production and income.

Following the close relationship between the economy and the polity found in these earlier thinkers, Adam Smith is sometimes regarded as separating economics from religion and politics (see Minowitz 1993), and building an analytical system in which the laws of motion of the economy are examined by treating the economy as a separate entity, separate from the polity and society (see Walter 2011). It is claimed that Smith achieved this by seeing the individual as being self-interested in the sense of being interested in furthering his or her material conditions, and by prescribing that political forces through government policies be kept to a minimum to allow the invisible hand of the market to make the social product as high as possible based on individual self-interest. While there are strands in Smith's writings that support elements of this interpretation, it is in many respects an inaccurate exaggeration. He took the view that free markets driven by individual self-interest alone do not always benefit society, and favoured some amount of government intervention and restrictions on individual freedom (Kurz 2015). He acknowledged that the interests of merchants and manufacturers were opposed to those of the rest of society, since they had an inherent tendency to deceive and oppress others, and thus emphasized the need for sympathy toward others and not just self-interest. He also argued that growing specialization had its downside, making workers who were engaged in repetitive tasks ignorant and dissatisfied. To counter this, he recommended the expansion of education financed by the government.

Ricardo is seen as pushing the separation of the economy from society and politics even further, and to see the functioning of the economy as a self-contained system in which the laws of production, accumulation and distribution can be seen as being dependant on the forces of diminishing returns in agriculture, the accumulation behavior of capitalists, and the law of population which, at least in the long run, determined wages. While in terms of his simple analytical structure, Ricardo's theory of wages was determined at subsistence by population dynamics, he argued that "[i]t is not to be understood that the natural price of labour, estimated even in food and necessaries, is absolutely constant. It varies at different time in the same country, and very materially differs in different countries. It essentially depends on the customs and habits of the people" (Ricardo 1817, 96–7). Distribution, thus, depends on social and political factors, and approach which was developed later by Marx.

John Stuart Mill stated that "The laws and conditions of the Production of wealth partake of the character of physical truths. There is nothing optional or arbitrary

in them. … It is not so with the Distribution of wealth. That is a matter of human institution solely" (1848, p. 199). He argued that "[i]n so far as the economical condition of nations turns upon the state of physical knowledge, it is a subject for the physical sciences, and the arts founded on them. But in so far as the causes are moral or psychological, dependent on institutions and social relations, or on the principles of human nature, their investigation belongs not to physical, but to moral and social science, and is the object of what is called Political Economy" (Mill 1844, 21; see also Smith 1985).

In Marx's analysis of the economic dynamics, the interaction between economic, political and social factors is made explicit in his method of dialectical materialism. The base of his system refers to the forces and social relations of production, and the superstructure contains its culture, political power structures, laws, social norms, and the state, and although the base is the starting point of his analysis, changes in the base affect elements in the superstructure and they, in turn, affect elements in the base, including technology and the nature of conflict between classes, in a reciprocal way. Marx's economic analysis of capitalism examines how production and income are distributed between capitalists—who own the means of production—and workers who work in capitalist enterprises with a wage at a low level due to the existence of a reserve army of the unemployed, but which is affected by political and social factors, what Marx described as moral and historical forces. Capitalist income in the form of profits is largely saved and invested, leading to capital accumulation and the expansion of output and employment. This expansion has resulted in enormous growth of production, and the success of capitalism led to a particular type of state, education and law, which tended to keep capitalist power in place. However, capitalism was prone to crises, for instance, due to mechanization, which was likely to lead to a falling rate of profit. This contradiction within capitalism, according to Marx, was likely to result in social and political struggle which spelt the death knell of capitalism and the emergence of socialism.

While Marx went in one direction, the marginalists, Leon Walras, Stanley Jevons and Carl Menger, among others, went in another. For instance, Walras's (1874) book created a self-contained economic general equilibrium system with supply and demand functions for different commodities and markets, the latter being derived from the utility obtained by individuals, without reference to the morality or immorality of doing so. The marginalists also developed the marginal product theory of distribution, in which profit maximizing, perfectly competitive firms demand factors of production, and in which factor prices clear markets with each factor price being equal to its value of marginal product. Given individual endowments, technological conditions that determine the marginal product relations, determine distribution. The separation of economics from political and social factors was complete, although some, like Marshall, were somewhat recalcitrant on this issue.

But subsequent mainstream or neoclassical economics went further in the direction of separateness. The notion of economics as the subject that examines how scarce resources are allocated to satisfy human wants can be, and sometimes is, examined for one person, sometimes called Robinson Crusoe (with Friday usually forgotten), who does not even live in society, but solves the economic problem. When

many individuals exist, and engage in production activity, they do so in firms that only involve technological relations but no social relations, and when they trade they do so in market exchange not requiring political and social involvement. Atomistic agents have no power, and conditions of power do not enter this economic analysis. When power enters at all, it does so in terms of market power, in which monopolists, oligopolists and monopolistic competitors have some market power, and can set prices given consumer demand conditions and cost considerations.

These views of the economy are reflected in the definitions of economics that are widely used in textbooks and other works. Many such definitions follow Lionel Robbins (1932, p. 15) who, as noted earlier, defined economics as "the science which studies human behaviour as a relationship between ends and scarce means which have alternative uses." As Backhouse and Medema (2009) document, variants of this definition came to be used in many textbooks although it was a tortuous path. For instance, Stigler (1942: 12) defined economics as "the study of the principles governing the allocation of scarce means among competing ends when the objective of the allocation is to maximize the attainment of the ends," and McConnell (1969: 23) stated that "Recalling that wants are unlimited and resources are scarce, economics can be defined as the social science concerned with the problem of using or administering scarce resources (the means of producing) so as to attain the greatest or maximum fulfillment of society's unlimited wants." This practice continues even today. For instance, Mankiw and Taylor (2011: 2) define it as "the study of how society manages its scarce resources."

The definition was met with criticisms from various quarters. Quite early on Knight (1951, 6) argued that the Robbins definition neglected the "liberal" view of life, and the system of "capitalism," since scarce resources can be allocated to meet ends by a single individual—hence the Robinson Crusoe analogy—and by a command economy with central planning, as Samuelson's textbook noted explicitly. As Groenwegen (1987) notes, the definition also was irreconcilable with Keynes's theory of unemployment, since in it, labor was not a scarce resource. Moreover, the definition does not, at least directly, address the issue of the distribution of income and wealth.

Not all definitions follow it, however. In the text which popularized the term economics in 1890, Marshall defined economics (and political economy) as "a study of mankind in the ordinary business of life; it examines that part of individual and social action which is most closely connected with the attainment and with the use of material requisites of well-being. Thus it is on the one side a study of wealth and on the other and more important side, a part of the study of man" (Marshall 1890, 1). Marshall did not specify exactly what he meant by the "ordinary business of life," but from other parts of the book, it is clear that he refers to "consumption and production, the distribution and exchange of wealth; … wholesale and retail dealing; foreign trade, and the relations between employers and the employed…" (Marshall 1890, 33).

Textbooks do not invariably use the Robbins definition. Paul Samuelson in his extremely popular introductory economics text (Samuelson 1948) which has gone into 19 editions at the time of this writing, As mentioned earlier, he defines the eco-

nomic system as one which answers the following questions about the production of goods and services: what, how and for whom. However, Samuelson's later editions, coauthored with William Nordhaus, defines economics as "the study of how societies use scarce resources to produce valuable goods and services and distribute them among different individuals" Samuelson and Nordhaus 2010: 4), although still maintaining the what, how and for whom questions (Samuelson and Nordhaus 2010: 7–8). Krugman and Wells (2009: 2) define it as "the social science that studies the production, distribution, and consumption of goods and services," in the introductory chapter called "the ordinary business of life" harking back to Marshall (1890).

The problems mentioned about the wants-scarcity definition of economics, however, have not affected the way the economy and economics are represented in textbooks. The basic tool used is the market demand–supply model, also returning to Marshall's (1890, 290) twin blades of the scissors, representing demand and supply. Thus, the economy is seen as a decentralized market economy which allocates scarce resources using the price mechanism, which provides signals to economic agents and rations goods and services among buyers. The distribution of income is determined in this world with endowments and factor prices according to what has been called marginal productivity theory. This is initially interpreted in terms of perfect markets with perfect competition (without which, strictly speaking textbook supply curves do not exist), but then monopolies, monopolistic competition and oligopolies are introduced using the profit maximizing framework, and market failures due to things such as asymmetric information, externalities and public goods. Initially the scarcity definition with full utilization of resources did not fit well with the macroeconomics portion of textbooks, which generally showed how income and output are determined not with supply and demand curves but using the income expenditure or diagonal cross model. However, through time, even macroeconomics came to be presented in terms of aggregate demand and aggregate supply curves, by analogy with market demand and supply curves, and unemployment, if it is seen to exist (as it is not in classical, new classical and real business cycle approaches), is explained by wage–price rigidity, despite the problems often associated with this tool (see Dutt 2002).

After the advent of marginalist economics, neoclassical economics (which organizes thinking on the basis of the optimizing individual or organization) steadily came to dominate economics, and became mainstream economics (although the latter also includes those who use sophisticated econometric methods). However, there were other schools of thought that resisted this trend at different times. The historical approach had a strong presence in the German region before the rise of marginalism and continued to hold some sway, taking a broader view of the economy. In the USA, institutionalist economists such as Thorstein Veblen and John Commons resisted the deductive and narrow approach, emphasizing the importance of shared social habits of the mind and laws that regulate behavior. In India Ranade (1898) also stressed institutional issues such as property rights in land and social norms that militate against entrepreneurship, in addition to the policies of the colonial government that perpetuated poverty. And the followers of Marx, including those who examined the issue of imperialism, took into account the interaction of economic, political and social factors. The economics of aggregate demand as developed by Keynes (1936),

who focused on how decision-making under uncertainty involved social and other institutional factors, and Kalecki (1971), who emphasized market and class power in determining distribution and the power of capitalists in preventing the pursuit of full employment polices by the state also provided a major alternative.

As noted earlier, there are some notable examples of neoclassical economists who use the individual optimizing framework to broaden their interests and go beyond the boundaries of economics, narrowly defined. Some of them extend the rational choice framework beyond economics, to political and social spheres. Although these can be seen as bringing together the study of the economy, polity and society, they can also be seen as narrowing the scope of what is brought into the study of politics and society by taking self-interest to be the only motivator and extending the market metaphor. Some of them have to do with bring in transaction costs and the analysis of institutions, establishing the rules of the game in terms of "good" institutions that strengthen property rights and make markets operate with less frictions and more flexibility, continuing to promote free market economics. Although this kind of economics as a prescriptive approach does not follow logically from the examination of markets with optimizing behavior—witness the many examples of market failure that result in inefficiency and the importance that may be attached to goals other than efficiency—but the tendency from self-seeking individuals to overcome a few market failures such as externalities by virtue of the Coase theorem and the ingenuity of people, and to argue against issues of fairness either by excluding value judgements in the name of science or, alternatively, the dismissal of inequality as a form of injustice by relating it to the morally problematic notion of envy—has had the effect of establishing a connection between neoclassical economics and the support for relatively free markets. Finally, behavioral economists, who have now entered mainstream economics, have questioned whether individuals act "rationally" and in a self-interested manner and have opened up room for social norms and state intervention to nudge people to improve their well-being (Thaler 2015). However, by still maintaining an individualistic approach, where individuals can improve their situation by learning how to overcome their biases (Tversky and Kahneman 1981), and to the extent that these biases can be "explained" by adding constraints to the optimizing approach (see, for instance, Rabin 2013), it is unclear how this approach will actually broaden economics.

The broad conclusion of this section is that from its early beginnings when the economy was studied with politics, society and ethics, there has been a significant narrowing of how the economy has been studied, reaching its zenith (or, perhaps, its nadir) with the rise to dominance of neoclassical economics. This narrowing has not gone uncontested by heterodox scholars, recently even within the mainstream of economics. However, it is not clear whether this has really been a broadening of mainstream economists, and how deep that broadening is.

6 The Problem with Specialization

The narrowing of economics that makes it focus only on the economy is problematic for at least three considerations (and the same can be said for other disciplines, for instance, politics and its focus on the polity). To examine them, we may refer to economics—following our earlier discussion—as being the study of activities such as producing, consuming, buying, selling, investing and holding wealth or the study of the system—the economy—which decides on what is produced, how and for whom, and how this changes over time. Analogously, we may define politics as the study of governing, government policy making, voting for political leaders, and waging war involving state and other organizations, and the study of the system, including the state, in which collective decisions affect the entire society are made and of how collective decision-making bodies—such as the government—are selected; sociology can be described as the study of all social behavior and relations not covered by economics and politics, and how the society operates and changes.

The first consideration refers to the fact that phenomena and institutions in the economy interact with those in the polity and society (in a narrow sense) as defined in the introduction Thus, the activities of buying and selling in markets (usually considered to be what happens in the economy) can have effects on social relations by weakening (if people involved in exchange see themselves as being antagonistic) or strengthening social ties (if coming together builds social bonds), and what happens in social relations—for instance, a decline in trust—can have an effect on how markets operate by adversely affecting whether contracts will be honored. Also, a decline in production and income in the economy can adversely affect government revenues and government expenditures on basic government services, reducing the political legitimacy of the state, and in turn, political unrest can increase uncertainty, reduce investment and income and production. These interactions, of course, are well-known, but they imply that not taking them into account by confining attention to only the "economy" results in incomplete and inadequate analyses.

The second, arguably more fundamental, consideration is that many—perhaps, most—concepts, organizations and behaviors cannot be considered to be strictly in the economy (or the political or social spheres), so that many aspects of the economy, society and polity are co-constituted, rather than simply interacting while being separate. Thus, markets have an economic dimension in which people and organizations buy and sell, a political dimension in which the state enforces law and order and regulates, and a social dimension in which individuals interact in ways beyond simply exchanging goods. Moreover, the state has an economic dimension as a producer and allocator of goods and services, a political dimension because it governs and exerts its power over society, and a social dimension in how people within the state and its elements interact with each other and with the rest of society. Consumption has an economic dimension in which people buy and use goods and services, a social dimension in which they consume in ways to convey meaning to others and which develops their own identity, and even a political dimension to express their support for political ideas and to influence political decisions. Production is not simply a

technical and economic activity, but a social one in which people possibly cooperate and build social communities, and even a site of political struggle between workers and capitalists.

It may be noted that the closed economic neoclassical system in which self-interested individual optimizing agents operate in self-organizing markets without taking into consideration "non-economic" aspects of society is problematic in its own terms. Why, if individuals optimize, why should they not seize other people's property and violate contracts to gain at the expense of others? The simple neoclassical formulation ignores the problems of the enforcement of property rights and contracts. These problems raise the importance of social norms and trust (see, for instance, Greif 2006) which govern and shape the behavior of individuals and groups and prevent them from always resorting to theft and the violation of contracts, and laws, police and the judicial system, which create a room for the state and politics. These problems are sometimes addressed within what can broadly be called neoclassical economics using transaction cost and institutional analysis in which institutions try to reduce such costs, but do not take adequate account of different sources of power and how they affect institutions and behavior.

Finally, if social inquiry involves understanding society with a view to examining what—if anything—is wrong and what can be done to overcome them, and even to change society, there needs to be some idea of what is good, or is an improvement, for society, and therefore some notion of ethics. Even if the goal is simply to understand what is happening, the questions that are been asked and what is focused on implicitly rely on value judgments; as is increasingly being understood, as noted earlier. What is good for society involves various aspects of society (in a broader sense), as is obvious whether one takes a utility perspective (since utility can depend on individual consumption, the consumption of others, time spends with friends and family, and political structures and policies),[5] a freedom and rights perspective (since these involve political, social and economic aspects, especially when what are called positive freedoms and rights are included), and a functionings and capabilities perspective (since what is valued by individuals and groups involves not just material consumption but also social and political aspects of life). To focus on one dimension, for instance, the production and income approach and its fixation with economic issues, is to take an incomplete and misleading perspective (see, for instance, Sen 1991; Stiglitz et al. 2010). Without an understanding of different aspects of society (in a broad sense) it is clearly not possible to understand or to change society.

[5]See, for instance, for the influence of economic, social and political issues, as determinants of life satisfaction, subjective well-being or self-reported subjective happiness, a standard way of measuring utility, see Layard (2006) and Radcliff (2013).

7 Why the Shift from Political Economy to Economics?

If we accept the definition of political economy as the field of study which examines the interaction and the co-constitution of what are referred to as economic, political and social issues (among other related ones), we may try to examine why political economy gave way to the more specialized discipline of economics.

First, with the expansion of knowledge about the economy (and the same is true for other objects of inquiry), scholars cannot be knowledgeable about all the accumulated knowledge about it. They will therefore need to select some relevant area of knowledge and know it well, and thus be competent in an area, rather than having a superficial knowledge of many things. The argument here is that since one cannot be an expert on economics, politics, sociology, and ethics, among other things, it is beneficial for a person to specialize, and be an expert on some things rather than a dilettante in many. The result of this, at can be claimed, will not only be a deeper understanding of what is studied, but also improve the process of knowledge accumulation, since we have true experts creating this specialized knowledge and learning by doing from this process. If scholars need to learn about matters beyond their expertise, all they need to do is to learn it from scholars in these areas. Moreover, scholars in a particular field will also benefit from having greater interaction with other scholars in their field, to know their own field better.

However, as discussed in the previous section, it is not at all clear that specialization of the particular type that has taken place, that is, between economics, politics and society, and in subfields within them, actually improves the process of knowledge creation. The quality of knowledge creation on a particular issue in general requires knowledge input from within a particular discipline as well as knowledge from other disciplines, as well as the cross-fertilization of knowledge on different disciplines. If the actual pattern of specialization takes place haphazardly without an assessment of the effects of this specialization (which leads to the omission of knowledge from other disciplines and their cross-fertilization), there is no particular reason why knowledge creation on a particular issue, and therefore knowledge creation overall, will be improved by this pattern of specialization.[6] While micro-motives may well exist for promoting specialization, the overall consequences of this may well be negative, if the co-constitution and interaction between economic, social, political and other concepts and issues is widespread. The problem may not be overcome simply by collaborating with people from other disciplines by forming "teams" especially

[6]It may be instructive to compare the process of division of labor and specialization in production and in society more generally, and specialization in the "production" of knowledge. Although the overall benefits of specialization have been extolled by early writers such as Plato (for the social division of labor) and Adam Smith (for the division of labor in the production of particular goods), there are many reasons to doubt that the benefits always exist, as recognized even by Smith (1776), who discusses the problem that workers who are engaged in repetitive tasks become ignorant and dissatisfied. The production of knowledge, moreover, is different from the production of goods and services, since the latter is concerned in large part with the quantity of production (for instance, increasing dexterity allows people to produce more, and increasing mechanization also raises productivity) while the former is concerned mainly with quality and not quantity.

if the interaction/constitution issues are underappreciated by those specializing in different disciplines, although it may be somewhat improved upon. The problem of needing to know far too many things may be overcome by choosing specialization on issues, such as inequality and consumption, rather than in established disciplines and sub-disciplines.

Second, there are sociological reasons why specialization according to fields takes place, reasons that are connected with the professionalization of fields. Marshall, who we noted, was the first to prominently use the term economics, is also considered to be a pioneer in the professionalization of economics (see Maloney 1985). Maloney focuses on three criteria of professionalization emphasized by sociologists: first, training in a specialized body of theory; second, an effective monopolization of a defined specialist function or functions; and third, observation of professional ethics maintained by colleagues' sanctions or force of opinion. Training in a specialized body requires establishing a well-defined curriculum of study that requires wide acceptance, and one that is likely to be hierarchical in nature, that is, there needs to be core knowledge that everyone in the profession needs to possess, and there may be fields of specialization that build up on the core. The core and specialization needs to be capable of being taught in the classroom although practical training and apprenticeships may be necessary is some, mainly practical, fields. Effective monopolization is obtained by limiting access to education, by making obtaining credentials difficult, by developing jargon and methods that are not widely understood, and by dismissing those who have not been properly "educated" as amateurs, mere journalists, and even charlatans. The creation and evolution of the content of the field according to which it is decided what is included in it and what is not, require leadership from people who have prestige, and who "produce" students and other followers who increase their power and control over the profession, and organizations, including academic departments, think tanks, journals and professional organizations, all with their own hierarchies. In turn, these leaders and organizations serve as arbiters of quality within the profession, as gatekeepers, and as preservers of professional standards and ethics (though regarding the latter there is relatively little; see DeMartino 2011).

Third, although there may be social pressures within the economics profession for specialization, what explains the specific pattern of specialization? When an academic profession effectively monopolizes a field of study, how does it select which aspects of study to monopolize? While there are no clear logical answers to this question, three related issues should be recognized, first, that the profession should make a strong case for why ut can claim to be experts at the study of some things or ideas, second, that these claims are made in relation to claims made by other professions regarding their own turf and third, that they can keep other people, such as policy makers and journalists, outside. The most obvious way to specialize is by focusing on activities and aspects of society in which issues of the production of goods and services, and market exchanges are central, and to do so in ways in that minimize influences outside these spheres, that is, making it self-contained, by holding them to be "given," or sometimes by totally ignoring them, as being outside their sphere of expertise, which can be left to other professions. The exclusion of others occurs,

of course, with professional training, and also by making the material relatively inaccessible to others, by using jargon and techniques that are difficult to understand, let alone master. Making acquiring knowledge sequentially—learning some things, and sequentially building on it—is the usual way of doing this. In some ways this is achieved by making disciplines more like mathematics or the natural sciences, adopting mathematical methods, and theories involving "first principles" and then building on them. In economics, the individual optimizer has served as a useful way of starting, as has the formalization of "pure" microeconomics and macroeconomics, and the study of econometrics. This formalization is also linked with keeping a narrow focus on "economic" issues, and emphasizing the importance of theory rather than policy issues, and deriving policies from theoretical constructs, all of which make it easier to leave out "non-economic" issues which may be more difficult to formalize mathematically because they involve complex phenomena, and because policy making involves complex issues well beyond narrow theoretical economics. All this is not done in a planned and coordinated way, but the result of many small actions by many scholars, though the writing of textbooks and journal articles, and by collective action by departments in universities and by professional organizations, serving to enforce quality and professional standards.

A skeletal view of some of the central features of mainstream economics can illustrate the nature of this specialization. Production of goods and services takes place in firms which have production functions which reflect given technology that convert inputs into output and sell them in perfectly competitive markets. Individual agents own inputs, including their labor, have given preferences over bundles of goods and services, and sell inputs and buy goods and services in perfectly competitive markets as price takers. Markets clear through price changes, and this system determines all (relative) prices, levels of production, and the incomes of all agents. It can be seen that in this skeletal system is self-contained and it is so because it leaves out of consideration aspects of society that involve things that can be studied in other disciplines. Thus, psychological and social issues are excluded by taking individuals as having given preferences which usually involve only goods and services they consume, or their incomes, without going into where the preferences come from and without introducing the effects of the consumption of others. Social and political issues are ignored by assuming markets exist, with private property rights and contracts are automatically enforced. Also, social and political issues in production and labor markets are ignored by making production depend on technological factors and making the labor market like any other market.

Of course there are countertendencies. First, some of the things that are ignored or exogenous in the skeletal construction can be brought in or endogenized at a later state. However, these extensions are path dependent, affected by the initial approach, since there is an effort to not change the approach too much so as to shake its very foundations. Second, phenomena that are not considered narrowly "economic" can be brought into the realm of analysis, like marriage, childbirth and politics. Economists, because of their so-called rational choice approach, and their sophisticated mathematical tools, have found it relatively easy to trespass into other disciplines. Third, other disciplines adopt the methods and approaches of economics,

blurring boundaries between disciplines based on these grounds. But this serves to enhance the prestige of economics.

Fourth, there are vested interests that support specialization in specific ways from outside the academy which influence how the scholars in the field view their own subject. The most obvious issue here is the attempt by those whose interests lie in "freeing" markets from what they consider undue political and social influences. Groups and classes in society that are engaged in business activities and who are in high income groups can see their interests lying in keeping government restrictions on behavior through regulations and direct competition in production activities, to a minimum, and to preventing redistributive government policies and social welfare policies from reducing their income. The market versus state debate can be seen in this light, with those who want to limit state activity in the economy arguing in favor of self-regulating markets through the mechanism of the invisible hand, without state intervention in markets. These same groups are also likely to see the encroachment of social factors that can possibly interfere with their market freedoms, through the influence of social norms and perhaps more importantly, through the actions of social groups, as a threat to their quest for increasing their income and power due to the existence of some types of social norms and social action. The tendency to change reality which separates the economy from the polity and from society, creating a separate social entity called the "economy" was discussed by Polanyi (1944) in terms of his analysis of the great transformation of societies that separated the economy from politics and society. This separation did not always reduce the role of the state in the economy, but in fact the state was used to forcibly separate the economy from politics and society.

How better to achieve their ends than by supporting a discipline which seeks to minimize or obliterate political and social issues, providing intellectual support for the world they want to create? This can be achieved by funding the research of scholars and groups of scholars who are willing to support this kind of separation in various ways, as by funding universities and research institutes and influencing the kind of teaching and research they do.

A few additional comments are necessary by way of clarification, First, the process is not planned collectively by some groups or classes, but results from a shared view—which may called an ideology—which seems to match their individual and collective perceived interests. Second, the process is influenced by the ideas of dominant members of the economics profession, who may have views that are similar but possibly independent of the views of those outside the academy, not least to maintain their power and prestige, and given that they are reasonably rich. Third, this process is not always in the actual interests of the groups that may support it—for instance, some of business interests may well want the state to control the economy in which it will benefit them, by providing them with special privileges, or by making the economy do well to enable them to have more buoyant markets in which they can sell more. Although these ideas are often recognized by members of these groups, they are underplayed, both because they may not be able to benefit from an overall change in the discipline and may possibly be harmed by them, and except under exceptional circumstances (for instance, in a sharp downturn in income and

employment), support the approach and get what benefits they can in specific ways without challenging the status quo. Finally, the pressure to separate the economy from the polity did not always come in the same manner. While in more recent times, politics is separated in an attempt to free market participants from restrictions which may be set on market behavior by the state which, in many democracies, attempts to protect the economically weak, but who are large in number and may want to regulate economic activity, in earlier times, for instance, in the time of Adam Smith, the separation was wanted at least to some extent to prevent the state from being used by powerful merchants and business to increase their power, possibly against the general good. As noted earlier, Smith was quite aware that the interests of merchants and manufacturers were opposed to those of the rest of society, since they had an inherent tendency to deceive and oppress others.

8 Conclusion

This paper has been concerned with the change in the name of the study of the economy from political economy to economics, and has argued for the replacement of economics by political economy.

It has examined the meanings of the two terms, political economy and economics. Economics and political economy have been distinguished for being a science rather than an art. This distinction, in turn, has a number of implied, but not identical distinctions. There is, first a distinction between scientific (or systematic theoretical and empirical analysis) of how the economy is as opposed to policy prescription about what should be done. There is something to this distinction, since actual policy advice and policymaking needs more than theoretical analysis or even a study of the facts relating to the theory—it needs wisdom, experience and "feel" for the relevant areas of the economy. However, the view of political economy supported here does not concern this distinction and is not even the same as Mill's distinction between science and art. Political economy in our sense needs to have systematic analysis based on theory and empirical knowledge, and this analysis can even lead to broad ideas about policies about what should be done. However, there also needs to be the art of political economy, since the systematic analysis will have to be coupled with wisdom and feel for particular contextual situations which can inform, but not replace, actual policymaking and advice. Second, there is the distinction between positive versus normative analysis, where the former involves the world as it is, whereas the latter involves making value judgments about what is ethically desirable. While there may positive and normative *questions*, the distinction between positive and normative *analysis* has been convincingly debunked in view of the fact that what is selected as being important for the analysis and how the analysis is formulated and expressed involves implicit value judgments, and also because purely "positive" conclusions cannot be definitively drawn on objective, scientific grounds because of epistemological issues concerning hypothesis testing (see, for instance, Dutt and Wilber (2013). Thus, political economy analysis necessarily involves value judge-

ments and normative issues. Finally, there is a distinction between pure economics and applied economics or political economy, where there something that is "pure" in the sense of being abstract theory or uncontaminated by "reality" which can then be "applied" to particular areas of the economy, to specific contexts and to policymaking. Since theory in the sense of political economy necessarily takes into account important aspects of reality taking into account the purposes to which the theory is to be used, there is no distinction between pure and applied analysis except in the sense of "organizing principles" which provide general guides to analysis without being related to reality (and hence is necessarily incomplete as theory) and in the sense of practical local wisdom that cannot be systematically analyzed (as noted in the discussion of policymaking earlier).

Our chosen definition of political economy refers to the broadening of the analysis of the economy to factors that can be considered economic, social, political, psychological and ethical, among others. It does not advocate any particular approach to the study of political economy, but it opens up room for different approaches, including Marxian and radical, post-Keynesian, institutionalist, structuralist, feminist, ecological and other approaches (on this, see also Rothschild, 1989). In this sense political economy can be associated with heterodox approaches to economics. However, it also includes neoclassical approaches to broadening, although not in the form which systematically excludes other approaches and methods or which insists only on the application "economic" methods—whatever that may mean, to the study of political behavior and institutions. It does not see itself as being different from economics, and in this sense it can seen as being another name for that field as interpreted by Groenwegen (1987), but it does see "economics" and "political economy" as being similar in content, and calls for significant change, which can be acknowledged by the return to the name "political economy"

The paper has argued the change in name did not coincide exactly with the change in the nature of the discipline, and the latter can be seen as a long drawn-out process of narrowing the scope of the study of the economy. This narrowing has created problems for the study of the economy because of the co-constitution and interrelations between the objects of study by the different disciplines and because of the fact that the well-being of people and societies depends on multiple factors that have been examined in different disciplines. Moreover, this narrowing has largely been the result of influences that have sought to promote the vested interests of powerful people and groups, often to the detriment of the general good of society and the relatively poor, excluded and marginalized. While political economy has witnessed a revival, it is often seen as being different from and sometimes a branch of economics. It is time to replace economics by political economy in name and, more importantly, in substance.

References

Aristotle (350 BCE). *Politics*, tr. By T. A. Sinclair. Harmondsworth, UK: Penguin Books, 1962.

Arndt, H. W. (1984a). Political economy. *Economic Record*, *60*(170), 266–273.

Arndt, H. W. (1984b). Political economy: A reply. *Economic Record*, *61*(175), 752.

Arndt, H. W. (1991). The purposes of political economy. *Methodus*, *3*(2), 113–116.

Backhouse, R. E., & Medema, S. G. (2009). Defining economics: The long road to acceptance of the Robbins definition. *Economica, 76,* 805–820.

Baran, P. A. (1957). *The political economy of growth*. Harmondsworth: Penguin Books, 1973.

Boesche, R. (2002). *The first great political realist: Kautilya and his Arthashastra*. Lanham: Lexington Books.

Campbell, W. F. (1987). Commentary: The old art of political economy. In S. T. Lowry (Ed.), *Pre-classical economic thought* (pp. 31–42). Boston: Kluwer Academic Publishers.

Caporaso, J., & Levine, D. (1992). *Theories of political economy*. Cambridge: Cambridge University Press.

Cohen, B. J. (2008). *International political economy: An intellectual history*. Princeton, New Jersey: Princeton University Press.

DeMartino, G. F. (2011). *The economist's oath: On the need for and content of professional economic ethics*. Oxford: Oxford University Press.

Dutt, A. K. (2002). Aggregate demand-aggregate supply analysis: A history. *History of Political Economy, 34*(2), 321–363.

Dutt, A. K., & Wilber, C. (2013). *Economics and ethics. An introduction*, revised and expanded edition. London and New York: Palgrave Macmillan.

Eaton, J. (1966). *Political economy: A marxist textbook*. New York: International Publishers Company.

Edgeworth, F. Y. (1925). *Papers relating to political economy* (Vol. 3). London: Macmillan.

Fraser, L. M. (1937). Economic thought and language: A critique of some fundamental concepts. London: A. C. Black.

Fusfeld, D. R. (1972). *Economics*. Lexington, MA: D C Heath.

Gilpin, R. (1975). *US power and the multinational corporations: The political economy of foreign direct investment*. New York: Basic Books.

Gilpin, R. (1987). *The political economy of international relations*. Princeton, NJ: Princeton University Press.

Greif, A. (2006). *Institutions and the path to the modern economy: Lessons from medieval trade*. Cambridge: Cambridge University Press.

Groenewegen, P. D. (1985). Professor Arndt on political economy: A comment. *Economic Record, 61*(175), 744–751.

Groenwegen, P. D. (1987). Political economy and economics. In J. Eatwell, M. Milgate & P. Newman (Eds.), *The new palgrave: A dictionary of economics* (Vol. 3, 904–907). London: Macmillan.

Groenwegen, P. D. (1992). *What is political economy*? *Methodus*.

Jevons, W. S. (1871). *Theory of political economy* (1st ed.), Preface. Included in 4th edn. London: Macmillan, 1910.

Jevons, W. S. (1879). *Theory of political economy* (1st ed.), Preface. Included in 4th edn. London Macmillan.

Kalecki, M. (1971). *Selected essays on the dynamics of the capitalist economy*. Cambridge University Press.

Kautilya. (1965). *Arthashastra*, ed. and tr. by R. P. Kangle. Delhi: Motilal Banarasidass.

Keynes, J. M. (1936). *The general theory of employment, interest and money*. London: Macmillan.

King, J. E. (1948). The origin of the term political economy. *Journal of Modern History, 20*(3), 230–231.

Knight, F. H. (1951). *Economics*, *on the history and method of economics*. Chicago: University of Chicago Press, 1963.

Krugman, P., & Wells, Robin. (2009). *Economics* (2nd ed.). New York: Worth.

Kurz, H. (2015). Adam Smith on markets, competition and violations of natural liberty. *Cambridge Journal of Economics, 39,* 1–24.
Layard, R. (2006). *Happiness. Lessons from a new science*. The Penguin Press.
Lowry, S. T. (1987). The Greek heritage in economic thought. In S. T. Lowry, (Ed.). *Pre-classical economic thought* (pp. 7–30). Boston: Kluwer Academic Publishers.
Maclcod, H. D. (1878). *Economics for beginners*. London: Longmans, Green and Co.
Maloney, J. (1985). *Marshall, orthodoxy and the professionalisation of economics*. Cambridge: Cambridge University Press.
Malthus, T. R. (1820). *Principles of political economy*. London: John Murray.
Mankiw, G., & Taylor, M. P. (2011). *Eonomics* (2nd ed.). Andover, UK: South Western.
Marshall, A. (1890). *Principles of economics* (8th ed.). London: Macmillan, 1920.
Marshall, A. (1961). *Principles of economics* (9th variorum ed.) C. W. Gullbaud, (Ed.). London: Macmillan.
Marshall, A., & Marshall, M. P. (1879). *Economics of industry*. London: Macmillan.
Menger, C. (1871). *Grunds tze de Volkswirtchaftelehre*. In C. Menger, Gesammelte Werke, Vol. 1 (2nd ed.). T bingen: J. C. B. Mohr.
McConnell, C. R. (1969). *Economics: Principles, problems and policies*. New York: McGraw-Hill.
Mill, J. S. (1844). *Principles of political economy with some of their applications to social philosophy*. London: Longman, Green, Reader and Dyor, reprinted, Clinton, NJ: Augustus Kelley, 1974.
Milonakis, D., & Fine, B. (2009). *From political economy to economics: Method, the social and the historical in the evolution of economic theory*. Abingdon, UK and New York: Routledge.
Minowitz, P. (1993). *Profits, priests, and princes. Adam Smith's emancipation of economics from politics and religion*. Stanford, CA: Stanford University Press.
Montchretien de A. (l615). *Traicte de L'Oeconomie Politique*, avec notes et un introduction par T. Funck-Brentano, Paris, 1899.
North, Douglass C. (1991). Institutions. *Journal of Economic Perspectives, 5*(1), 97–112.
Pareto, V. (1927). *Manual of political economy,* translated by A. S. Schwier, M. Kelley. New York, August 1968.
Phelps, E. (1985). *Political economy: An introductory text*. New York: W W Norton.
Polanyi, K. (1944). *The great transformation*. Boston: Beacon Press, 1957.
Rabin, M. (2013). Incorporating limited rationality into economics. *Journal of Economic Literature, 51*(2), 528–543.
Radcliff, B. (2013). *Political economy of human happiness*. Cambridge and New York: Cambridge University Press.
Ranade, M. G. (1898). *Essays on Indian economics. A collection of essays and speeches.* Madras: G.A. Natesan and Co. (2nd ed., 1906).
Ricardo, D. (1817). *On the principles of political economy of taxation*. In P. Sraffa (ed.). (1951). *The works and correspondence of David Ricardo* (Vol. 1) Cambridge: Cambridge University Press. (3rd ed., 1821).
Robbins, L. (1932). *An essay on the nature and significance of economic science* (2nd ed.). 1945. London: Macmillan.
Robbins, L. (1937). *The economic basis of class conflict and other essays in political economy*. London: Macmillan.
Robbins, L. (1981). Economic theory and political economy. *American Economic Review*, 71(2), 1–10.
Robinson, J., & Eatwell, J. (1973). *An introduction to modern economics*. New York: McGraw-Hill.
Rothschild, K. W. (1989). Political economy or economics? Some terminological and normative considerations. *European Journal of Political Economy, 5*(1), 1–12.
Samuelson, P. A. (1948). *Economics*. New York: McGraw-Hill.
Samuelson, P. A., & Nordhaus, W. D. (2010). *Economics* (19th ed.). Boston and New York: McGraw-Hill and Irwin.
Schumpeter, J. A. (1954). *History of economic analysis*. Oxford: Oxford University Press.

Sen, A. (1991). *Development as freedom*. New York: Knopf.
Sidgwick, H. (1883). *Principles of political economy.* London: Macmillan.
Smith, A. (1776). *An inquiry into the nature and causes of the wealth of nations*, London: Methuen, reprinted in *The Glasgow edition of the works and correspondence of Adam Smith* (Vols. 1 and 2). Oxford: Oxford University Press.
Smith, V. R. (1985). John Stuart Mill's famous distinction between production and distribution. *Economics and Philosophy, 1,* 267–284.
Staniland, M. (1985). *What is political economy? A study of social theory and underdevelopment*. New Haven and London: Yale University Press.
Steuart, Sir J. (1767). *Inquiry into to principles of political oeconomy*. London: Millar and Cadell.
Stigler, G. J. (1942). *The theory of competitive price*. New York: Macmillan.
Stiglitz, J., Sen, A., & Fittousi, J.-P. (2010). *Mismeasuring our lives: Why GDP doesn't add up*. New York and London: The New Press.
Stilwell, F. (2002). *Political economy: The contest of economic ideas*. Oxford and New York: Oxford University Press.
Sturtevant, J. M. (1877). *Economics or the science of wealth*. New York: Putnam.
Thaler, R. (2015). *Misbehaving: The making of behavioral economics*. New York: W W Norton and Co.
Trautmann, T. R. (2012). *Arthashastra. The science of wealth*. London: Penguin Books.
Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science, 211,* 453–458.
Verri, P. (1763). Memorie storiche sulla economia pubblica dello sato di Milano. In *Scrittori Classici Italiani di Economia Politica*, Parte Moderna, Vol. XVII, Milan, 1804.
Viner, J. (1948). Power versus plenty as objectives of foreign policy in the seventeenth and eighteenth centuries. *World Politics*, 1(1), 1–29.
Walras, L. (1954). *Elements of pure economics*, translated by William Jaffe, London: Allen & Unwin.
Walter, R. (2011). *A critical history of the economy*. Abingdon, UK: Routledge.
Weingast, B. R., & Wittman, D. A. (Eds.). (2006a). *The oxford handbook of political economy*. Oxford: Oxford University Press.
Weingast, B. R., & Wittman, D. A. (2006b). The reach of political economy, in Weingast and Wittman, eds. (2006a)
Whynes, D. (Ed.). (1984). *What is political economy? Eight perspectives*. Oxford: Blackwell.

Public–Private Partnerships, Corruption and Inefficiency

Indrani Roy Chowdhury and Prabal Roy Chowdhury

1 Introduction

Governments around the world struggle to provide better services to their citizens on limited budgets. The search for new methods for the production and delivery of public services has given us new concepts such as alternative service delivery (ASD) mechanisms, and more specifically public–private partnerships (PPPs). ASD refers to the full sets of alternative arrangement of public goods and service provisioning and delivery, that would otherwise have been provided directly by the public enterprises alone. This will include PPPs, but will also include contracting out of services and outright privatization. By contrast, a PPP is a long-term contractual arrangement between a government and a private party for the provisioning of assets, and delivery of services that have been traditionally provided by the public sector. It also includes sharing of risk and rewards among the private and the public entities (Iossa and Martimort 2015).

PPPs have been widely used in Europe, Canada and the USA as part of a general trend. Even in the developing countries, PPP arrangements have grown steadily since the 1990s. According to the World Bank's Private Participation in Infrastructure (PPI) database, during 1990–2003, about 1000 projects and 47% of investments in Latin American and Caribbean countries involved the use of PPPs. In the central and the

This chapter draws on the authous'earlier work, Roy Chowdhury and Roy Chowdhury (2016).

I. R. Chowdhury
Jawaharlal Nehru University, New Delhi 110067, India
e-mail: indraniroychowdhury01@gmail.com

P. R. Chowdhury (✉)
Indian Statistical Institute, Delhi Center, Katwaria Sarai, 7 - S.J.S.S. Marg, New Delhi 10016, Delhi, India
e-mail: prabalrc1@gmail.com

P. Ray et al. (eds.), *Economics, Management and Sustainability*,
https://doi.org/10.1007/978-981-13-1894-8_4

eastern European countries, PPPs started mushrooming in the second half of the 1990s.

While public–private partnerships (PPPs) have only caught the imagination over the last few decades, interestingly India has a long history in this respect. In fact India could boast of PPPs as early as the nineteenth century, e.g. the Great Indian Peninsular Railway Company (1853) and the Bombay Tramway Company (1874). In India PPPs really took off around the 1990s though, with the period between 2006–2011 seeing a spectacular growth in the number of PPPs from 450 projects in November 2009 (worth Rs. 2242 billion), to 758 PPP projects (worth Rs. 3833 billions in July 2011). Considering the sector-wise distribution by values, in July 2011, roads constituted 53.4%, urban development 20.1%, ports 8%, tourism 6.6% and airports 0.7% of all PPPs.

Distribution across states was also skewed, with top five states accounting for 58.3% of the total value of PPPs. These include Andhra Pradesh, Maharashtra, Karnataka, Gujarat and UP, with these states focusing on roads, ports and airports.[1,2] In India, various forms of PPPs have been employed. These include turnkey contracts,[3] Build Operate Transfer (BOT) contracts,[4] performance-based management / maintenance contracts,[5] etc.

The Indira Gandhi International Airport in New Delhi, a partnership between the government of India and the GMR group (one of top five airport developers), is of course the one undisputed success story in this respect. Its terminal 3 (T3) was built in an impressive thirty-seven months. It is the only airport in India with three runways, one being India's longest at 4430 metres. The T3 can also boast of 168 checkout counters, 92 wakalators, 31 escalators, 78 aerobridges and 12 baggage reclaim belts, handling 20.6 million domestic and 9.3 million international passengers in the financial year 2011. Further, it was rated world's fourth best airport for *airport service quality* (*ASQ*) in 15–25 million passenger category in FY 2011. Other examples of successful PPPs include the Rajiv Gandhi International Airport in Hyderabad.

A cautionary tale is, however, provided by the PPPs tasked with upgrading the London tube. In 1993, this project was awarded to two different infra-firms, Tube-

[1]Much of the preceding discussion draws on the PPP database from the Department of External Affairs, Ministry of Finance.

[2]In Andhra Pradesh prominent projects include the HITEC City, Hyderabad, RGI Airport, and the Krishnapatnam Port. Projects in the pipeline include Hyderabad metro rail project, bridge across Godavari at Rajahmundry, Machilipatnam port. The RGI airport, Hyderabad, partnered with the GMR group, handled 5.8 mn. domestic and 1.9 mn. international passengers in the financial year 2011. It boasts of the second longest runway in India, 146 check-in counters, and 46 immigration counters. It was rated the world's best airport for Airport Service Quality (ASQ) in the 5–15 million passenger capacity in the financial year 2011.

[3]These are beneficial in terms of time and cost savings, as well as risk sharing, since payments and penalties can be linked to performance.

[4]BOT contracts can take various forms. These include (a) User fee-based BOTs, practiced in roads, ports airports, and energy, (b) Annuity-based BOTs, not meant for cost recovery through user charges (practiced in rural urban health education sectors)

[5]Encouraged in the sectors which are constrained in terms of economic resources, e.g. water supply, sanitation, road maintenance, etc.

lines and Metronet, after 19 months of contract negotiations (see the Appendix for some more details). Metronet ran into significant problems however. By April 2005, Metronet had only started work on 13 out of 32 scheduled stations and was also 12 months behind on the refurbishment of 78 district lines.[6] Moreover, it received only 121 million pounds out of the 551 million needed to cover its cost overruns. Metronet went into administration between July 2007 and May 2008, after which it was bought by two Transport for London (TFL) subsidiaries. The remaining one-third of the PPP, being run by Tubelines, was taken back into public control in May 2010 after seven and a half years for a purchase price of 310 m. While Tubelines did not have overspend problems, there were delays.[7]

Not surprisingly therefore, people have argued both in favour of, as well as against PPPs. While Manmohan Singh, our ex-Prime Minister, said that 'PPP projects ..... will not only enable us to leverage our limited public resources, but also improve efficiency of service delivery', the New Zealand Treasury was much more circumspect 'There is little reliable empirical evidence about the costs and benefits of PPPs'.

The evidence regarding the performance of PPPs is actually mixed. A report commissioned by the Treasury Taskforce estimated a savings of 17% on PPPs in UK, as compared to public procurements. NAO (2003) reported that innovative designs helped reduce the level of staffing and thus led to an overall cost reduction of 30% in UK. Private finance initiative (PFI) projects in UK seems to be delivering costs savings compared to the traditional procurement. Improvement in completion time and costs of delivery has been achieved and 76% of the PPP projects have been completed on time as compared to 30% of the traditionally procured projects (Andersen 2002). However, Blanc-Brude et al. Blanc-Brude (2009) studied a sample of road projects in all EU-15 countries plus Norway. They found that ex ante construction costs are 20% higher for PPP roads. Similarly, PPPs have resulted in higher water prices than traditional procurement in France and seem to be unsuitable for fast moving sectors in UK.

In this paper, we want to focus on one possible cost associated with PPP formation, namely the possibility of collusion and corruption among the various PPP partners. This seems a relevant concern since PPPs open up the scope of pork-barrel politics based on ideology, social or political ties, or simply incentive to pander. Levin and Tadelis (2010) document that local political institutions have a profound impact on such choices in the USA. In France also, it seems efficiency considerations are often secondary to government's determination to provide private benefits to some groups, with delayed costs. Further, the fact that PPPs typically involve longer procurement periods and tend to be complex, so that there are few consortiums capable of undertaking them (four per project on the average in the UK), increases the chances of collusion.

[6] In the first year's deductions amounted to 32 million, and bonuses were only about 12 million. Despite the delays Metronet had significant profits at higher than the market average rate, suggesting the penalties were not too large.

[7] See, among others, Iossa and Martimort (2016).

We first develop a simple analytical framework that extends Chowdhury and Chowdhury (2016). Chowdhury and Chowdhury (2016) builds on the consensus in the literature to argue that one major advantage of PPPs, among others, is that of risk sharing.[8] Given that most projects undertaken via PPPs are reasonably large, various risks, including demand shocks are a serious concern, and consequently risk sharing becomes important. They argue that depending on whether governance is weak or strong, PPPs may or may not form. We extend their framework to allow for the possibility of collusion and corruption among the various partners of a PPP. We then show that PPPs are more likely to form in case the externality gains are significant, and the agents are quite risk averse. Otherwise, the government may opt for government control instead since PPP formation may lead to bribery and sub-optimal project choice.

There is a small but developing literature on the role of corruption in PPPs. Martimort and Poyet (2008) examine a scenario where there is collusion between contractors and government officials at an early stage where the government is deciding whether to form a PPP, or opt for traditional procurement. While Iossa and Martimort (2016) examine the role of collusion within a PPP, where collusion ensures that contracts between the contractors and the government is left incomplete, allowing transfer of surplus to the contractor. This possibility in turn means that contracts must take simple forms so as to reduce collusive possibilities, which in turn implies that risk sharing is inefficiently low. In the present paper also collusion if any takes place after the formation of a PPP, as in Iossa and Martimort (2016). Unlike Iossa and Martimort (2016), however, collusion implies that the project may be scrapped even if it is socially efficient for the project to go through. Further, the focus is on the relative merits of PPPs vis-á-vis other organizational forms, and how collusive possibilities affect these choices.

In the next section, we discuss the formal model, before turning to the analysis in Sects. 3 and 4. Finally Sect. 5 concludes.

2 A Formal Framework

We develop a formal model consisting of several agents, the government, a government department (denoted G) and a private firm (denoted F). Potentially the government department and the private firm can form a public–private partnership, denoted PPP. There is a single public project which might be taken up by either G or F, or even the PPP in case it forms. It is the government which decides whether it wants to use the governmental department to implement the project. In case the

[8]There are other reasons of course, most notably synergy between the private and government sectors.

government decides, otherwise, it can either invite the private sector or alternatively try and form a PPP to implement the project.[9]

The project yields a benefit of X to those sections of the population directly affected by it, and an indirect benefit of P to the rest of the economy, where $X, P > 0$. The direct benefit, X, can be extracted from the population by the project authority in the form of usage fees, etc. The indirect benefit, however, is in the nature of an externality and cannot be extracted. The project is brought to fruition in two stages, build and operate. Both the stages can be handled by all the agents, the private firm, the government department, as well as the PPP. For simplicity, we assume that the cost of each stage is the same for all three agents and for both stages, though it depends on the state of the world.[10] This per stage cost is c_3 if the state of the world is 'good', it is c_2 if the state of the world is 'indifferent', and c_3 if the state is 'bad', so that $c_1 > c_2 > c_3$. Let p_i denote the probability that the cost is c_i, where $0 < p_i < 1$ and $\sum_i p_i = 1$.

For ease of exposition, we assume that irrespective of the institutional organization of the project, the whole of the monetized value of the consumers' surplus is extracted and accrues to the organization implementing the project. While this is a serious assumption, especially if the project is run by either the government department, or the PPP, we feel that the analysis should qualitatively go through in a more realistic scenario. We assume that all agents can access 1 unit of capital at its opportunity cost of 1.[11]

The objective of the government is welfare maximization. The objective of the other agents will be described as we go along.

The timeline is as follows. We consider a one period two-stage game:

Stage 1. The government decides on the organizational structure, i.e. whether the project is to be implemented by the government department, the private firm, or the PPP.

Stage 2. The state of the world is revealed, with the agents in charge of project implementation getting to know the exact state of the world, while the government only gets to know if the state of the world is good or not. In case the state of the world is good, the project is necessarily implemented. Otherwise, the government department, if it is involved in project implementation, reports the state of world to the government and the project continues if and only if the report is indifferent. If the private firm implements the project, then the private firm reports to the government regarding

[9] Note that our formulation abstracts from strategic interactions in project allocation. While lobbying for such projects is undoubtedly an important concern, the present formulation allows us to focus on the issue of potential corruption post PPP formation and not confound it with any possible corruption earlier.

[10] We should flag that this is a serious assumption in that one of the primary reasons behind PPP formation is actually absolute advantages that the government department and the private firm may enjoy, creating a synergy in case of PPP formation. Again our justification is that this allows us to focus on the issue of potential corruption in PPPs without confounding issues.

[11] It is possible that the cost of accessing capital is different for the government department and the private firm. However, we abstract from it for simplicity.

the state of the world, and the project goes through iff the report is indifferent. We assume that information regarding whether the state is indifferent or bad is soft, so that misreporting is possible. This potential misreporting creates a possibility of bribery and corruption under PPP formation, something that we elaborate on later.

We solve for the subgame perfect Nash equilibrium of this game.

3 The Analysis

Let $u(.)$ denote the government, as well as the government department's utility function, where $u(.)$ satisfies the usual conditions, i.e. $u(0) = 0,\ u'(.) > 0,\ u''(.) < 0$. Note that $u(.)$ is concave, which shall play an important role in the analysis. The following assumption allows us to focus on the case of interest.

Assumption 1 (a) $P + u(X - 2c_2) > 0 > X - 2c_2$.
(b) $0 > P + 2u(X/2 - c_1)$.
(c) $P + u(X - 2c_3), X - 2c_3 > 0$.

Assumption 1(a) states that if the state of the world is indifferent, i.e. the cost of project implementation is c_2, then doing the project is efficient, but yields a negative profit. Whereas Assumption 1(b) states that if the state of the world is bad, in that the cost of project implementation is c_1, then doing the project is inefficient. In fact, from Assumption 1(a), it yields a negative profit as well. Finally, Assumption 1(c) ensures that in the good state of the world doing the project is efficient and also yields a positive profit. Note that one implication of this assumption is that the interests of the government department, and the private firm is really opposed when the state of the world is indifferent, and it is this fact, as we shall see later, creates a possibility of corruption later on.

As is usual, we start by solving the second stage game first, before solving for the optimal organizational structure.

3.1 Government Department Does the Project

The *government department* cares about the social externality P arising out of the project, the income accruing to the department (call it g), as well bribery income, if any (call it B). Formally, the utility of the government department is given by

$$P + u(g) + B. \tag{1}$$

In case the state of the world is good, i.e. the cost of project implementation is c_3, the project is implemented. This follows since from Assumption 1(c), both the

government and the department want to implement the project which is efficient. In that case, the utility of the government department is $P + u(X - 2c_3) > 0$.

If the state of the world is indifferent, i.e. c_2, then the department reports truthfully to the government, and the project is implemented. This follows since given Assumption 1(a), implementing the project is efficient, so that the departmental utility is $P + u(X - 2c_2) > 0$.

Whereas, from Assumption 1(b), the project is not implemented in case the state of the world is bad. This follows since $0 > P + 2u(X/2 - c_2) > P + u(X - 2c_2)$ from the concavity of $u(.)$. Thus, the expected utility of the department is

$$P(p_2 + p_3) + \sum_{i=2}^{3} p_i u(X - 2c_i). \tag{2}$$

Note that this is also the social welfare since by assumption the whole of the consumers' surplus is extracted. This is one of the expositional advantages of our assumption that whichever firm does the project extracts the whole of the surplus from the consumers.

Proposition 1 *In case the project is implemented by the government department, it reports the state of the world truthfully under every eventuality, and the project is implemented iff the state of the world is either good, or indifferent. The expected utility of the department, as well as social welfare, is*

$$W^G = P(p_2 + p_3) + \sum_{i=2}^{3} p_i u(X - 2c_i). \tag{3}$$

Note that our result, that the department is necessarily truthful, is to some extent an artefact of this simple framework. Suppose that there are multiple states of the world such that it is efficient to implement the project under all such states, but the government cannot observe which one has occurred (though the department can), say c_2' and c_2'', where $c_2'' > c_2'$. Now, in case say c_2' occurs, then department may have an incentive to overstate the cost to be c_2'', and pocket the difference in costs $c_2'' - c_2'$. Note, however, that the department will implement a project iff it is efficient, so that allocational efficiency is not affected.

Finally, given Assumption 1(a) and 1(c), it is individually rational for the government department to agree to operate the project if offered.

3.2 Private Firm Does the Project

The utility of the private firm running the project, when it has a direct payoff of f, say, is given by

$$u(f). \tag{4}$$

It is straightforward to extend the analysis to the case where the utility function of the private firm is different from $u(.)$.

Given Assumption 1(a), note that $X < \min\{2c_2, 2c_1\}$, so that pursuing the project is unprofitable whenever the state of the world is either indifferent or bad. Thus in either case the private firm will report that the state of the world is bad, and the project is scrapped. The project goes through only in the good state of the world. Thus, the utility of the private firm, as well as social welfare, is given by

$$p_3 u(X - 2c_3). \tag{5}$$

Proposition 2 *Under a private firm, the project is implemented iff the state of the world is good. Irrespective of whether the state of the world is indifferent, or bad, the firm reports that it is bad and the project is scrapped. The expected utility of the private firm, as well as social welfare is given by*

$$W^F = p_3[P + u(X - 2c_3)]. \tag{6}$$

Note that given Assumptions 1(c), it is individually rational for the private firm to agree to run the project if offered.

3.3 Public–Private Partnerships

Under a PPP, the two participants agree to bear the responsibilities of investing in one of the stages. For concreteness suppose the private firm invests in building the project, whereas the government department takes care of the operational stage. Further, for simplicity, assume that the private firm and the government department each obtain half the surplus, i.e. $X/2$, in case the project is implemented.

If the state of the world is good, then the government observes it, and since it is efficient (Assumption 1(c)), the government ensures that the project is implemented. Whereas if the state of the world is bad, then neither the department, nor the private firm has an incentive to pursue the project (Assumptions 1(a) and 1(b)). Thus, the department will report the state of the world truthfully and the project will be scrapped.

Finally, suppose that the state of the world is indifferent, i.e. the per stage cost is c_2. Then while the private firm would prefer the project to be scrapped, it is efficient for the project to go through (Assumption 1(a)). This creates an incentive for bribery whereby the department reports the true state to be c_1, so that the project is scrapped in return for a bribe. Whether such a bribe happens or not depends on parameter values.

Note that the gain to the firm in case the project is scrapped is $c_2 - \frac{X}{2}$. Assuming that all the bargaining power is with the government department, it can demand

and obtain the whole of this amount as a bribe payment in case it misreports the state to be c_1, whereas the department's utility in case the project is not scrapped is $P + u(\frac{X}{2} - c_2)$. Thus, the department accepts the bribe and misreports iff

$$P + u(\frac{X}{2} - c_2) < c_2 - \frac{X}{2}. \tag{7}$$

Thus, the essential cost of bribery is that the indifferent project will not be implemented. The bribe itself is just a transfer and does not enter welfare calculations. It is of course reasonable to argue that it is ethically reprehensible.

Proposition 3 *Assume that* $\sum_{i=2}^{3} p_i u(\frac{X}{2} - c_i) \geq 0$.

(a) In case $P + u(\frac{X}{2} - c_2) < c_2 - \frac{X}{2}$*, the project goes through only in the good state. In the bad state, the project is scrapped with no bribe, whereas in the indifferent state the department misreports that the state is bad and the project is scrapped and the department gets a bribe of* $c_2 - \frac{X}{2}$*. Aggregate welfare is given by*

$$W^{PPP}(B) = p_3[P + 2u(\frac{X}{2} - c_3)]. \tag{8}$$

(b) In case $P + u(\frac{X}{2} - c_2) > c_2 - \frac{X}{2}$*, the project goes through in the good state, as well as the bad state, and the department always reports truthfully. Aggregate welfare is given by*

$$W^{PPP}(NB) = (p_2 + p_3)P + \sum_{i=2}^{3} 2u(\frac{X}{2} - c_i)p_i. \tag{9}$$

Note that given Assumption 1, and the proposition hypothesis, it is individually rational for both the government department and the private firm to agree to the PPP contract.

4 The Government's Organizational Decision in Stage 1

We then solve the whole game, showing that optimally the PPP is going to form whenever there is no bribery under PPP formation. Otherwise, the optimal organizational design may lead to governmental control.

There are several trade-offs at play here. In the absence of bribery, the welfare gains under PPP vis-á-vis government operation follows from risk sharing among the government and the private firm. There is an additional welfare gain from PPP formation vis-á-vis private operation since under a PPP the project will necessarily be implemented even if the state is indifferent. In the presence of bribery, however,

PPP formation may involve additional costs as indifferent but efficient projects may be scrapped by the government department in return for bribes.

Straightforward calculations yield

Proposition 4 *(a) Welfare under government ownership exceeds that under private ownership.*

(b) In case $P + u(\frac{X}{2} - c_2) > c_2 - \frac{X}{2}$, *welfare under PPP exceeds that under either government, or private ownership.*

(c) In case $P + u(\frac{X}{2} - c_2) < c_2 - \frac{X}{2}$, *the welfare comparison between a PPP and government ownership is ambiguous. However, PPP formation dominates private ownership.*

Proof (a) This follows from a comparison of (3) and (6), since the indifferent project will be taken up under government ownership, but not under private ownership.

(b) From a comparison of (3) and (9), note that a sufficient condition for $W^{PPP}(NB)$ to exceed W^F is that $u(z/2) > u(z)/2$ for all z. This is true since $u(.)$ is concave. Finally, from part (a) of this proposition, it follows that $W^G > W^F$.

(c) The comparison between PPP and government ownership involves several trade-offs; while PPP improves risk sharing, it also leads to scrap the project in case the state of the world is indifferent. Depending on which of these effects dominate, the comparison may go either way.

Finally, PPP formation dominates private ownership since PPP formation improves risk sharing.

Can one find conditions such that the government necessarily prefers governmental control over PPP formation? One set of sufficient condition is that the externality gains P is not too large, so that one has that $P + u(\frac{X}{2} - c_2) < c_2 - \frac{X}{2}$, which which ensures that under the indifferent state there will be bribery, leading to the project being scrapped. If further the agents are risk neutral, then the risk diversion benefits of PPP formation will vanish. Thus in this case, the government will prefer government control over PPP formation.

We then consider the consequences of relaxing some of the assumptions.

1. Suppose that under a PPP, the private firm obtains $(1 - \alpha)X$ and the government obtains αX, where $0 < \alpha < 1$ is different from $1/2$. Note that under (i) an appropriately modified version of A1 and (ii) $p_3[u(X(1 - \alpha) - c_3] > 0$ (so as the ensure that the private firm accepts the contract), it is straightforward to check that given risk aversion all the propositions still go through.
2. In case the utility function of the government $u(P + g)$, instead of $P + u(g)$ as in the text. Again the results go through qualitatively, as they are essentially driven by the fact that the government and the private firm are risk-averse, rather than the specific functional forms adopted here.

5 Conclusion

Given the volume of infrastructural investment required, PPPs seem essential for India's development. We analyse a phenomenon that has been relatively under-analysed in the literature, namely the possibility of collusion between the private firm and the governmental department. We first develop a simple formal model based on risk sharing that is capable of analysing this issue. We then show that PPPs are more likely to form in case the project has significant externality gains, and the agents are quite risk averse. Otherwise, PPP formation may lead to bribery and sub-optimal project choice, and the government may opt for government control instead.

Appendix: The London Metro PPP

Tubelines were awarded the Jubillee, the Northern and the Picadelly lines, while Metronet got (a) the Bakerloo, Central and Victoria, and (b) the District, Circle, Hammersmith and City lines. The negotiations were extremely complex, with the work being phased over 30 years. Tubelines and Metronet were to do the upgradation, while the London Underground Limited (LUL) was to provide the final services. In fact, the complexity of the contract meant that advisory services themselves cost 109 million pounds. Further, the delay in project allocation meant that the LUL had to provide substantial compensations to the various bidders, including the losers.

Broadly, the contract specified two aspects, risk allocation among the parties, and the payment mechanism. As to risk allocation, the demand risk was borne by LUL as they did the final service provision. The cost overrun risks were, however, shared, as the amount of cost overruns to be borne by the private companies were capped. This reduced the incentives for checking overruns, as is clear from the Metronet experience. Metronet's incentive for checking costs overruns was not very strong, as Metronet's shareholders were also suppliers of the consortium, so that any higher costs would translate into revenue for them! In fact Metronet had cost overruns four times that of expected costs. However, Tubelines which did not have such an interlinked structure incurred no cost overruns. Finally, financing risk was also shared. While financing was provided privately by the Infracos, the public sector-borne substantial risks via a debt guarantee entitling lenders to recoup 95% of their invested funds in case of early contract termination.

Turning to the payment mechanism, it involved basic Infrastructure Service Charge (ISC), combined with bonuses and deductions. Performance was measured in terms of journey time capability (JTC), i.e. the time needed for a train to complete a journey. Ambience and general conditions of trains were measured by consumer surveys.

References

Andersen, O. J. (2002). Public Private Partnership: A new channel for local mobilization and participation. In *ECPR Workshop: Institutional innovation and local democracy*.

Blanc-Brude, F., Goldsmith, H., & Välilä, T. (2009). A comparison of construction contract prices for traditionally procured roads and public–private partnerships. *Review of Industrial Organization*, *35*(1), 19–40.

Chowdhury, I. R., & Chowdhury, P. R. (2016). Public private partnerships and governance, in understanding development: An Indian perspective on legal and economic policy. In S. Banerjee, V. Mukherjee, S. K. Haldar, (Eds.), India: Springer.

Iossa, E., & Martimort, D. (2015). The simple micro-economics of public-private partnerships. *Journal of Public Economic Theory*, *17*, 4–48.

Iossa, E., & Martimort, D. (2016). Corruption in PPPs, incentives and contract incompleteness. *International Journal of Industrial Organization*, *44*, 85–100.

Levin, J., & Tadelis, S. (2010). Contracting for government services: Theory and evidence from U.S. cities. *The Journal of Industrial Economics*, *58*(3), 507–541.

Martimort, D., & Poyet, J. (2008). Build it or not: normative and positive theories of public-private partnerships. *International Journal of Industrial Organization*, *26*, 393–411.

NAO, National Audit Office. (2003). PFI Construction Performance. London: National Audit Office, HC371

Part II
Vulnerabilities and Inclusive Growth

Capitalism, Crisis and the Common Man

Chandana Ghosh and Ambar Ghosh

1 Introduction

When we were in the first year in Presidency College, Kolkata, Anupda introduced us to the history of the Soviet Union. He also referred to the book, "Soviet Economic Development since 1917" by Maurice Dobb (Dobb 1966). To our great pleasure, we found both his lectures and the book by Dobb equally gripping and fascinating. Soviet Union is no more. The world has become uni-polar. Capitalism and capitalists are ruling the world today. In this context, we gratefully acknowledge Anupda's contribution to our learning as regards how Soviet Union took care of its ordinary citizens from "cradle to grave" and, thereby, awakened hope among the poverty stricken toiling masses all across the world. In today's world, where hostility of the mighty capitalist powers gravely threatens the survival of the socialist states, we consider it important to write a paper on how global capitalists have made and are making the lives of common men miserable all across the globe.

Remarkably, following the collapse of the Soviet Union, the capitalist world is going through a prolonged period of recession. Japan is in recession since 1992. The USA and Europe are in recession since 2001 and 2008, respectively (see Table 1). In all these cases, recessions followed collapse of huge asset price bubbles. Stock and real estate price bubbles collapsed in Japan in 1991. A huge dotcom bubble crashed in the USA in 2001. In many European countries, real estate bubbles burst in 2008. In the USA, the recession that started in 2001 deepened into a severe crisis in the wake of a crash in a huge house price bubble. Speculative activities of global financial capital (global financial institutions) are at the root of all these troubles.

C. Ghosh
Indian Statistical Institute Calcutta, Kolkata, India
e-mail: chandanaghosh521@gmail.com

A. Ghosh (✉)
Jadavpur University, Kolkata, India
e-mail: ambarghosh@gmail.com

P. Ray et al. (eds.), *Economics, Management and Sustainability*,
https://doi.org/10.1007/978-981-13-1894-8_5

The latest victim is Greece, which plunged into a severe recession since 2008. The purpose of this paper is to explain how Greece plunged into a severe crisis since 2008, causing tremendous suffering of its common people, who lost jobs on a large scale and suffered drastic cuts in wages and welfare spending including retirement and unemployment benefits.

The dominant view explaining Greek crisis in the literature is the following: In the wake of the formation of Eurozone, there took place large inflows of capital from the centre of the Eurozone to Greece, as exchange rate risk disappeared. Borrowing cost as a result went down in Greece inducing the Greek government and private economic agents to borrow on a large scale. These borrowings made Greece's debt very large. This led to a sharp deterioration in the risk perception of the foreign investors regarding Greece and induced them to stop investing in its assets. With the drying up of foreign capital inflows, government and private spending declined substantially creating a severe recession in Greece. For a detailed exposition of this view, one may go through, for example, Gibson et al. (2014), Krugman (2013), Lane (2012), Dellas and Tavlas (2013). We shall subject this line of thought to close scrutiny in the light of the available evidences, identify the factors responsible for the Greek crisis and present our argument in a rigorous theoretical framework. A careful analysis of data reveals that the Greek crisis is on account of not just one but two factors, namely the recession in other European nations and the USA due to the collapse in the real estate bubbles and a large decline in capital inflows.

To achieve the goal stated above, we first develop a model, which is suitable for explaining the crisis in Greece. The standard IS-LM-based Keynesian macro-models have many deficiencies (for details, one may go though Jha et al. (2016) and Rakshit (1993)). The feature that makes these models unsuitable for explaining the Greek crisis is that they do not and cannot explicitly show how different kinds of expenditures are financed. In case of Greece, foreign capital inflows directly financed government and private expenditures. Obviously, this phenomenon cannot be captured within the frameworks of the models mentioned above. Another major shortcoming of the aforementioned models is that they do not consider financial intermediaries, which play a major role in financing investment and consumption expenditures. Even though Bernanke and Blinder (1988) incorporated banks in the IS-LM model, they could not show how bank loans finance different kinds of expenditures. We shall, therefore, for our purpose develop a model which incorporates financial intermediaries and explicitly shows how different types of expenditures are financed. Jha et al. (2016) develop such a model for an open economy without capital mobility. We shall extend the model to incorporate capital mobility and apply it to the case of Greece.

Table 1 Annual growth rate of GDP (Annual percentage growth rate of GDP at market prices based on constant local currency. Aggregates are based on constant 2005 US dollars)

	1981	82	83	84	85	86	87	88	89	90
China	5.2	9.0	10.8	15.2	13.6	8.9	11.7	11.3	4.2	3.9
France	1.1	2.5	1.3	1.5	1.6	2.4	2.6	4.7	4.4	2.9
Germany	0.5	−0.4	1.6	2.8	2.3	2.3	1.4	3.7	3.9	5.3
Greece	−1.6	−1.1	−1.1	2.0	2.5	0.5	−2.3	4.3	3.8	0.0
Ireland	3.3	2.3	−0.2	4.4	3.1	−0.4	4.7	5.2	5.8	8.5
Italy	0.8	0.4	1.2	3.2	2.8	2.9	3.2	4.2	3.4	2.0
Japan	4.2	3.4	3.1	4.5	6.3	2.8	4.1	7.1	5.4	5.6
KoreaRep (South)	7.4	8.3	12.2	9.9	7.5	12.2	12.3	11.7	6.8	9.3
Malaysia	6.9	5.9	6.3	7.8	−1.1	1.2	5.4	9.9	9.1	9.0
Portugal	1.6	2.1	−0.2	−1.9	2.8	4.1	6.4	7.5	6.4	4.0
Spain	−0.1	1.2	1.8	1.8	2.3	3.3	5.5	5.1	4.8	3.8
Thailand	5.9	5.4	5.6	5.8	4.6	5.5	9.5	13.3	12.2	11.2
UK	−0.8	2.1	4.2	2.3	3.5	3.2	5.5	5.9	2.5	0.5
USA	2.6	−1.9	4.6	7.3	4.2	3.5	3.5	4.2	3.7	1.9

(continued)

Table 1 (continued)

	1991	92	93	94	95	96	97	98	99	2000
China	9.3	14.3	13.9	13.1	11.0	9.9	9.2	7.9	7.6	8.4
France	1.0	1.6	−0.6	2.3	2.1	1.4	2.3	3.6	3.4	3.9
Germany	5.1	1.9	−1.0	2.5	1.7	0.8	1.8	2.0	2.0	3.0
Greece	3.1	0.7	−1.6	2.0	2.1	2.9	4.5	3.2	3.1	4.2
Ireland	1.9	3.3	2.7	5.8	9.6	9.3	11.2	8.9	10.8	10.2
Italy	1.5	0.8	−0.9	2.2	2.0	1.3	1.8	1.6	3.7	1.8
Japan	3.3	0.8	0.2	0.9	1.9	2.6	1.6	−2.0	−0.2	2.3
KoreaRep (South)	9.7	5.8	6.3	8.8	8.9	7.2	5.8	−5.7	10.7	8.8
Malaysia	9.5	8.9	9.9	9.2	9.8	10.0	7.3	7.4	6.1	8.9
Portugal	4.4	1.1	−2.0	1.0	4.3	3.5	4.4	4.8	3.9	3.8
Spain	2.5	0.9	−1.0	2.4	2.8	2.7	3.7	4.3	4.5	5.3
Thailand	8.6	8.1	8.3	8.0	8.1	5.7	−2.8	−7.6	4.6	4.5
UK	−1.2	0.4	2.6	4.0	4.9	2.7	3.1	3.4	3.1	3.8
USA	−0.1	3.6	2.7	4.0	2.7	3.58	4.5	4.4	4.7	4.1

(continued)

Table 1 (continued)

	2001	02	03	04	05	06	07	08	09	10
China	8.3	9.1	10.6	10.1	11.4	12.7	14.2	9.6	9.2	10.6
France	2.0	1.1	0.8	2.8	1.6	2.4	2.4	0.2	−2.9	2.0
Germany	1.7	0.0	−0.7	1.2	0.7	3.7	3.3	1.1	−5.6	4.1
Greece	4.1	3.9	5.8	5.1	0.6	5.7	3.3	−0.3	−4.3	−5.5
Ireland	5.8	5.9	3.8	4.4	6.3	6.3	5.5	−2.2	−5.6	0.4
Italy	1.8	0.3	0.2	1.6	0.9	2.0	1.5	−1.0	−5.5	1.7
Japan	0.4	0.3	1.7	2.4	1.3	1.7	2.2	−1.0	−5.5	4.7
KoreaRep (South)	4.5	7.4	2.9	4.9	3.9	5.2	5.5	2.8	0.7	6.5
Malaysia	0.5	5.4	5.8	6.8	5.3	5.6	6.3	4.8	−1.5	7.4
Portugal	1.9	0.8	−0.9	1.8	0.8	1.6	2.5	0.2	−3.0	1.9
Spain	4.0	2.9	3.2	3.2	3.7	4.2	3.8	1.1	−3.6	0.0
Thailand	3.4	6.1	7.2	6.3	4.2	5.0	5.4	1.7	−0.7	7.5
UK	2.8	2.5	3.3	2.5	3.0	2.7	2.6	−0.5	−4.2	1.5
USA	1.0	1.8	2.8	3.8	3.3	2.7	1.8	−0.3	−2.8	2.5

(continued)

Table 1 (continued)

	2011	12	13	14	15
China	9.5	7.8	7.7	7.3	
France	2.1	0.2	0.7	0.2	
Germany	3.7	0.4	0.3	1.6	
Greece	−9.1	−7.3	−3.2	0.7	
Ireland	2.6	0.2	1.4	5.2	
Italy	0.6	−2.8	−1.7	−0.4	
Japan	−0.5	1.8	1.6	−0.1	
KoreaRep (South)	3.7	2.3	2.9	3.3	
Malaysia	5.3	5.5	4.7	6.0	
Portugal	−1.8	−4.0	−1.1	0.9	
Spain	−1.0	−2.6	−1.7	1.4	
Thailand	0.8	7.3	2.8	0.9	
UK	2.0	1.2	2.2	2.9	
USA	1.6	2.3	2.2	2.4	

Source World Bank

2 The Model for an Open Economy with Imperfect Capital Mobility

We develop here a model for a small open economy with imperfect capital mobility. Following the Keynesian tradition, we assume that aggregate output is demand determined. Given this assumption, GDP is determined by the following equation:

$$Y = C(Y) + I(r) + \bar{G} + G(r) + \left[X\left(\frac{P^*e}{P}, Y^*\right) - M\left(\frac{P^*e}{P}, Y\right)\right] \tag{1}$$

In Eq. (1), $X \equiv$ exports, $M \equiv$ the value of imports in terms of domestic goods, $P^* \equiv$ the average price of foreign goods in foreign currency, $P \equiv$ the average price of domestic goods in domestic currency, $Y^* \equiv$ foreign GDP and $e \equiv$ nominal exchange rate. Government expenditure is decomposed into two components: one exogenously given component denoted $\bar{G}$ and another component $G(r)$, which is a decreasing function of interest rate. As the economy is small, P^* is given. P is taken to be fixed. Two exchange rate regimes are possible: the fixed and the flexible. Here, we focus only on the fixed exchange rate regime, as it is the one that is relevant for our purpose.

Fixed Exchange Rate Regime

The exchange rate is pegged at $\bar{e}$. Incorporating this pegged value of e into Eq. (1), we rewrite it as

$$Y = C(Y) + I(r) + \bar{G} + G(r) + \left[X\left(\frac{P^*\bar{e}}{P}, Y^*\right) - M\left(\frac{P^*\bar{e}}{P}, Y\right)\right] \tag{2}$$

Here, we incorporate cross-border capital flows and denote the net inflow of capital by K. For simplicity, we assume that K is exogenously given and denote its value by $\bar{K}$. Note that, here $\bar{K}$ is given in terms of domestic goods.

The central bank intervenes in the foreign exchange market to keep the exchange rate fixed at $\bar{e}$. $\left\{X\left(\frac{P^*\bar{e}}{P}; Y^*\right) - M\left(\frac{P^*\bar{e}}{P}, Y^*\right)\right\}\frac{P}{\bar{e}} + \bar{K}\frac{P}{\bar{e}}$ gives the excess supply of foreign currency at the given exchange rate. The central bank buys up this excess supply with domestic currency at the price $\bar{e}$ creating high-powered money to keep e at $\bar{e}$. We further assume for the purpose of illustration that the government borrows from the central bank to finance the autonomous component of its consumption expenditure. We assume that high-powered money is created only on account of government's borrowings from the central bank and central bank's intervention in the foreign exchange market to keep the exchange rate fixed. Thus, the increase in the stock of high-powered money in the period under consideration is given by

$$dH = P\bar{G} + \bar{e}\frac{P}{\bar{e}}\left[X\left(\frac{P^*\bar{e}}{P}, Y^*\right) - M\left(\frac{P^*\bar{e}}{P}, Y\right) + \bar{K}\right] \tag{3}$$

From Eq. (3) it follows that the stock of real balance created in the period under consideration is given by

$$\frac{dH}{P} = \bar{G} + \left[X\left(\frac{P^*\bar{e}}{P}, Y^*\right) - M\left(\frac{P^*\bar{e}}{P}, Y\right) + \bar{K} \right] \tag{4}$$

We assume that households do not take any loans, carry out all their transactions with bank deposits, hold all their wealth in the form of bank deposits and banks are the only source of loans to the firms. Foreign investors also invest their fund in bank deposits. These are all simplifying assumptions. We can easily incorporate other financial assets. Given the assumptions stated above, the whole of the high-powered money created will be held by the banks as reserve. Accordingly, the amount of new loans in real terms the banks will plan to supply to the firms in the given period, which we denote by l_f, is given by (see Eq. (4))

$$l_f = (1-\rho)\frac{\bar{G} + \left[X\left(\frac{P^*\bar{e}}{P}, Y^*\right) - M\left(\frac{P^*\bar{e}}{P}, Y\right) + \bar{K}\right]}{\rho} \tag{5}$$

where ρ denotes CRR. We ignore excess reserves for simplicity.

We have assumed in this paper that investors finance their investment entirely with bank loans, which is, by assumption, the only source of loans to the private sector. The government also finances a part of its expenditure with loans from commercial banks. Equilibrium in the loan market is, therefore, given by the following equation

$$(1-\rho)\frac{\bar{G} + \left[X\left(\frac{P^*\bar{e}}{P}, Y^*\right) - M\left(\frac{P^*\bar{e}}{P}, Y\right) + \bar{K}\right]}{\rho} = I(r) + G(r) \tag{6}$$

where $I(r)$ is the investment function of the firms. The specification of our model is now complete. It contains three key Eqs. (2), (4) and (6) in three unknowns Y, $\frac{dH}{P}$ and r. We can solve them as follows: We can solve Eqs. (2) and (6) for the equilibrium values of Y and r. Putting the equilibrium value of Y into Eq. (4), we get the equilibrium value of $\frac{dH}{P}$. We show the solution in Fig. 1, where in the upper panel the IS and LL schedules represent Eqs. (2) and (6), respectively, in the (Y, r) plane. The equilibrium values of Y and r correspond to the point of intersection of the IS and LL schedules. These equilibrium values of Y and r are labelled Y_0 and r_0, respectively. In the lower panel, where positive values of $\frac{dH}{P}$ are measured in the downward direction, the schedule HH represents Eq. (4). It gives corresponding to every Y the value of $\frac{dH}{P}$, as given by Eq. (4). The equilibrium value of $\frac{dH}{P}$ corresponds to the equilibrium value of Y on the HH schedule. We shall now illustrate the working of the model using a comparative static exercise.

Fiscal Policy: The Effect of an Increase in Government Expenditure Financed by Borrowing from the Central Bank

Suppose the government raises $\bar{G}$ and finances it by borrowing from the central bank. How will it affect Y, $\frac{dH}{P}$ and r? We shall examine this question first diagrammatically using Fig. 2, where the initial equilibrium values of Y, $\frac{dH}{P}$ and r are labelled Y_0, $\left(\frac{dH}{P}\right)_0$ and r_0, respectively. Y_0 and r_0 corresponds to the point of intersection of IS and LL schedules in the upper panel, while $\left(\frac{dH}{P}\right)_0$ corresponds to Y_0 on the HH

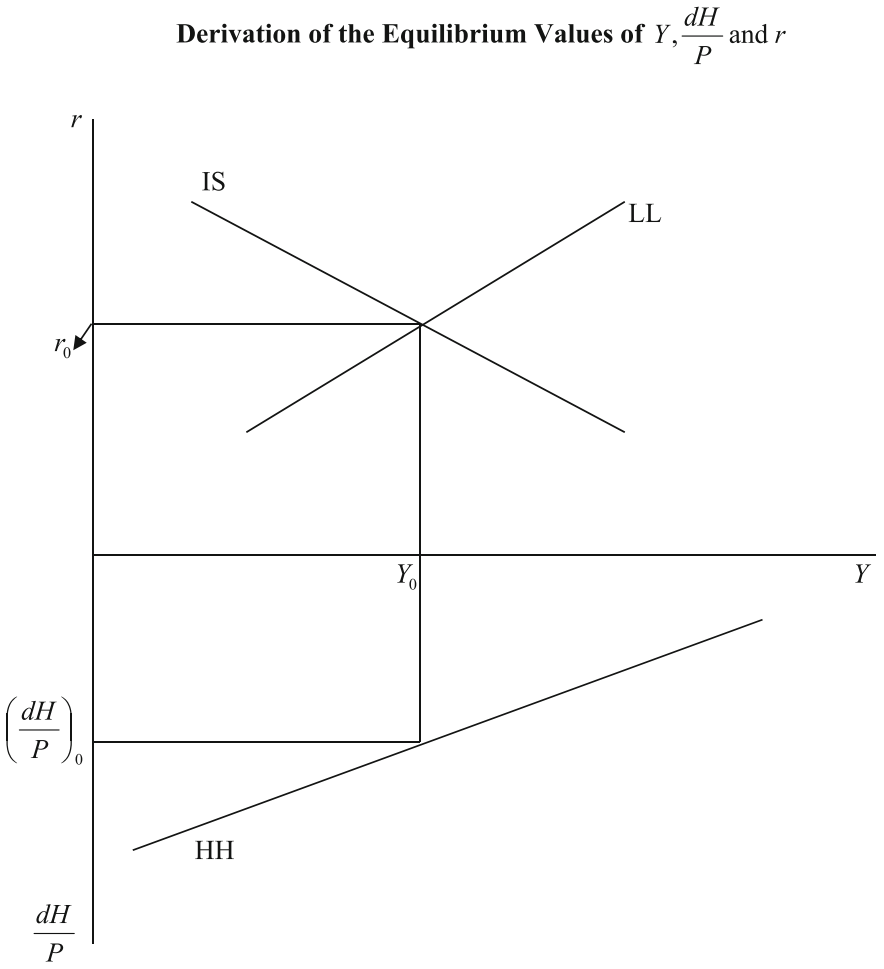


Fig. 1 Derivation of the equilibrium values of Y, $\frac{dH}{P}$ and r

schedule in the lower panel. First, focus on the IS curve. Take any (Y, r) on the initial IS. Following an increase in $\bar{G}$ by $d\bar{G}$ financed by borrowing from the central bank, there emerges an excess demand of $d\bar{G}$ for domestic goods at the given (Y, r). At the given r, therefore, the goods market will be in equilibrium at a larger Y, or at the given Y, the goods market will be in equilibrium at a higher r. Hence, the IS curve will shift upward or to the right. The new IS is labelled IS_1 in Fig. 2. Now, focus on the LL curve. Take any (Y, r) on the initial LL. Following the increase in $\bar{G}$ by $d\bar{G}$ financed by borrowing from the central bank, there now emerges at the given (Y, r) an increase in the supply of new loans by the banks—see the LHS of Eq. (6), while demand for new loans from banks as given by the RHS of Eq. (6) remains unaffected. Therefore, it follows from Eq. (6) that the loan market at the

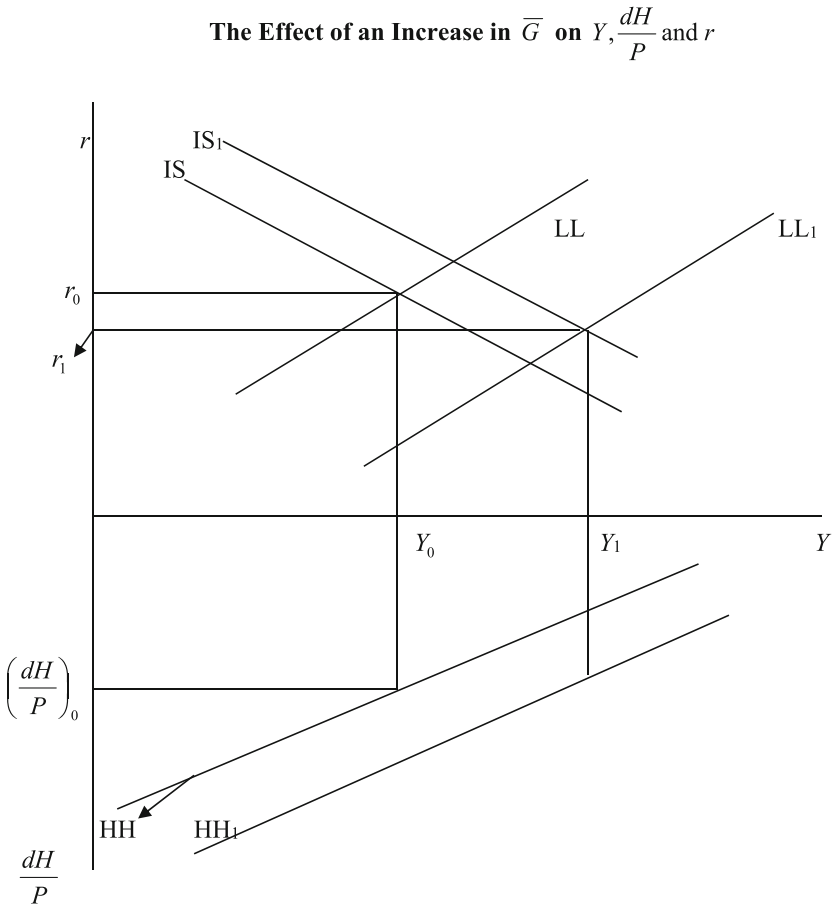


Fig. 2 The effect of an increase in $\bar{G}$ on $\bar{G}$

given r will be in equilibrium at a larger Y or at a lower r at the given Y. Thus, the LL shifts to the right or downward. The new LL is labelled LL_1. Hence, in the new equilibrium, Y will be larger unambiguously, but r may change in either direction. However, we have derived mathematically below that r will fall. Let us now focus on the HH schedule representing Eq. (4). Following an increase in $\bar{G}$ by $d\bar{G}$, supply of $\frac{H}{P}$, as given by the RHS of Eq. (4), increases by $d\bar{G}$ corresponding to every Y. Hence, the HH schedule will shift southward. The new HH schedule is labelled HH_1. Accordingly, the direction of change in the equilibrium value of $\left(\frac{dH}{P}\right)$ is ambiguous. However, we have mathematically derived below that it will increase.

Mathematical Derivation of the Results

To derive the results mathematically, we first substitute Eq. (6) into Eq. (2) to write it as

$$Y = C(Y) + \frac{1}{\rho}\left[\bar{G} + X\left(\frac{P^*\bar{e}}{P}; Y^*\right) - M\left(\frac{P^*\bar{e}}{P}; Y\right)\right] + \frac{1-\rho}{\rho}\bar{K} \tag{7}$$

Taking total differential of Eq. (7) treating all exogenous variables other than $\bar{G}$ as fixed, we have

$$dY = C'dY + \frac{1}{\rho}\left(d\bar{G} - M_Y dY\right)$$

Solving the above equation for dY, we get

$$dY = \frac{d\bar{G}}{\rho(1-C') + M_Y} \tag{8}$$

Again, taking total differential of Eq. (6) treating all exogenous variables other than $\bar{G}$ as fixed and using Eq. (4), we get

$$dr = \frac{\frac{1-\rho}{\rho}\left[d\bar{G} - M_Y dY\right]}{I' + G'} = \frac{\frac{1-\rho}{\rho}d\left(\frac{dH}{P}\right)}{I' + G'} = \frac{dl_f}{I' + G'} \tag{9}$$

Again, substituting Eq. (8) into the above equation, we get

$$dr = \frac{\frac{1-\rho}{\rho}\left[1 - \frac{M_Y}{\rho(1-C')+M_Y}\right]d\bar{G}}{I' + G'} < 0 \tag{10}$$

From Eqs. (4) and (8), we get

$$d\left(\frac{dH}{P}\right) = \left[1 - \frac{M_Y}{\rho(1-C') + M_Y}\right]d\bar{G} = \left[\frac{\rho(1-C')}{\rho(1-C') + M_Y}\right]d\bar{G} > 0 \tag{11}$$

Again, from Eqs. (5) and (11), we get

$$dl_f = \frac{1-\rho}{\rho}\left[\frac{\rho(1-C')}{\rho(1-C') + M_Y}\right]d\bar{G} = (1-\rho)\left[\frac{(1-C')}{\rho(1-C') + M_Y}\right]d\bar{G} \tag{12}$$

Adjustment Process

We shall now explain below how these changes come about. Following the increase in $\bar{G}$ by $d\bar{G}$, Y through the multiplier process increases by $\frac{d\bar{G}}{1-(C'-M_Y)}$. From this additional income, people save $(1-C')\frac{d\bar{G}}{1-(C'-M_Y)}$ and they hold this in the form of bank deposits. Banks receive an additional deposit of $(1-C')\frac{d\bar{G}}{1-(C'-M_Y)}$. Accordingly, their reserves and, therefore, the stock of high-powered money increase by $(1-C')\frac{d\bar{G}}{1-(C'-M_Y)}$. Let us explain this point a little more. When the government borrows from the central bank $d\bar{G}$ amount, the stock of high-powered money in the economy rises by the same amount. But following the increase in Y by $\frac{d\bar{G}}{1-(C'-M_Y)}$, import

demand rises by $M_Y\left[\frac{d\bar{G}}{1-(C'-M_Y)}\right]$ generating an excess demand for foreign currency (in terms of domestic goods) by the same amount. The central bank has to buy up $M_Y\left[\frac{d\bar{G}}{1-(C'-M_Y)}\right]$ amount of domestic currency (in terms of domestic goods) with foreign currency. Thus, at the end of the multiplier process the stock of high-powered money in the domestic economy rises by $d\bar{G}-M_Y\left[\frac{d\bar{G}}{1-(C'-M_Y)}\right]=\left(1-C'\right)\frac{d\bar{G}}{1-(C'-M_Y)}$. Banks get this, as we have already explained, in the form of additional deposits and reserve. Let us make this point clearer. As Y increases by $dY_1=\frac{d\bar{G}}{1-(C'-M_Y)}$, people's saving increases by $(1-C')\frac{d\bar{G}}{1-(C'-M_Y)}$. Besides this, they also have in their hands $M_y\frac{d\bar{G}}{1-(C'-M_Y)}$ part of their income, which they do not spend on domestic goods. Note that $(1-C')\left[\frac{d\bar{G}}{1-(C'-M_Y)}\right]+M_Y\frac{d\bar{G}}{1-(C'-M_Y)}=d\bar{G}$. However, they will not deposit $M_y\frac{d\bar{G}}{1-(C'-M_Y)}$ amount of income with the banks. They will sell it to the central bank for foreign currency. So, the banks will get an additional deposit of $(1-C')\frac{d\bar{G}}{1-(C'-M_Y)}$. In the central bank's balance sheet, the following changes will occur. On the asset side, central bank's credit to the government will increase by $d\bar{G}$ and its stock of foreign exchange will go down by $M_Y\frac{d\bar{G}}{1-(C'-M_Y)}$ so that, in the net, central bank's total asset increases by $d\bar{G}-M_Y\left[\frac{d\bar{G}}{1-(C'-M_Y)}\right]=\left(1-C'\right)\frac{d\bar{G}}{1-(C'-M_Y)}$. On the liabilities side banks' reserve rises by $(1-C')\frac{d\bar{G}}{1-(C'-M_Y)}$.

Banks will not want to keep the whole of the additional reserve idle. They will plan to extend an additional credit of $(1-\rho)\left(1-C'\right)\frac{d\bar{G}}{1-(C'-M_Y)}$. r will, therefore, fall by $\left[\left\{(1-\rho)\left(1-C'\right)\frac{d\bar{G}}{1-(C'-M_Y)}\right\}/\left(I'+G'\right)\right]$ to raise investment and government consumption by the amount of the additional supply of bank credit. This will bring about the second round of expansion in Y. At the end of the first round, increases in Y, dH and l_f and the decline in r are given, respectively, by $\frac{d\bar{G}}{1-(C'-M_Y)}$, $\left(1-C'\right)\frac{d\bar{G}}{1-(C'-M_Y)}$, $(1-\rho)\left(1-C'\right)\frac{d\bar{G}}{1-(C'-M_Y)}$ and $\left[\left\{(1-\rho)\left(1-C'\right)\frac{d\bar{G}}{1-(C'-M_Y)}\right\}/\left(I'+G'\right)\right]$.

In the second round, the increase in investment and government consumption by $(1-\rho)\left(1-C'\right)\frac{d\bar{G}}{1-(C'-M_Y)}$ will lead through the multiplier process to an increase in Y by $(1-\rho)\left(1-C'\right)\frac{d\bar{G}}{[1-(C'-M_Y)]^2}\equiv dY_2$. Out of this additional income of dY_2, people will save $(1-C')dY_2$ and will not spend $M_Y dY_2$ on domestic goods. Note that $(1-C')dY_2+M_Y dY_2=(1-\rho)(1-C')dY_1$, which is the amount of new credit extended by the banks at the end of the first round. However, the banks will not get back the whole of this credit as new deposit. People will deposit $(1-C')dY_2$ with the banks and sell $M_Y dY_2$ to the central bank. In the balance sheet of the central bank, following changes will occur. On the asset side, central banks' stock of foreign exchange will fall by $M_Y dY_2$, and on the liabilities side banks' reserve will go down by the same amount. In the second round, therefore, the stock of high-powered money will decline by $M_Y dY_2$. In the second round, aggregate saving increases by $(1-\rho)\left(1-C'\right)\frac{d\bar{G}}{[1-(C'-M_Y)]^2}$, which the households will hold in the form of bank

deposits. Banks will receive additional deposits of $(1-\rho)(1-C')^2\frac{d\bar{G}}{[1-(C'-M_Y)]^2}$, which will induce them to extend additional credit of $(1-\rho)^2(1-C')^2\frac{d\bar{G}}{[1-(C'-M_Y)]^2}$. This will increase investment and government consumption by the same amount through the decline in r by $\left[\left\{(1-\rho)^2(1-C')^2\frac{d\bar{G}}{[1-(C'-M_Y)]^2}\right\}/(I'+G')\right]$. Thus, another round of expansion will begin. This process will go on until the amount of additional investment and government consumption generated falls to zero. When that happens, the economy achieves a new equilibrium. Thus, the total increases in Y, dH, and dl_f and the decline in r are given, respectively, by

$$dY = \frac{d\bar{G}}{[1-(C'-M_Y)]} + (1-\rho)(1-C')\frac{d\bar{G}}{[1-(C'-M_Y)]^2} + (1-\rho)^2(1-C')^2\frac{d\bar{G}}{[1-(C'-M_Y)]^3} + \ldots = \frac{d\bar{G}}{\rho(1-C')+M_Y} \tag{13}$$

$$d\left(\frac{dH}{P}\right) = d\bar{G} - M_Y\frac{d\bar{G}}{[1-(C'-M_Y)]} - M_Y(1-\rho)(1-C')\frac{d\bar{G}}{[1-(C'-M_Y)]^2} - \ldots = \frac{\rho(1-C')d\bar{G}}{\rho(1-C')+M_Y} \tag{14}$$

$$dl_f = (1-\rho)(1-C')\frac{d\bar{G}}{[1-(C'-M_Y)]} + (1-\rho)^2(1-C')^2\frac{d\bar{G}}{[1-(C'-M_Y)]^2} + (1-\rho)^3(1-C')^3\frac{d\bar{G}}{[1-(C'-M_Y)]^3} + \cdots = \frac{(1-\rho)(1-C')d\bar{G}}{\rho(1-C')+M_Y} \tag{15}$$

$$dr = (1-\rho)(1-C')\frac{d\bar{G}}{[1-(C'-M_Y)]}\left(\frac{1}{I'+G'}\right) + (1-\rho)^2(1-C')^2\frac{d\bar{G}}{[1-(C'-M_Y)]^2}\left(\frac{1}{I'+G'}\right) + \ldots = \frac{(1-\rho)(1-C')d\bar{G}}{\rho(1-C')+M_Y}\left(\frac{1}{I'+G'}\right) \tag{16}$$

Clearly, (13)–(16) tally with the values of dY, $d(dH)$, dl_f and dr derived mathematically earlier and given by (8), (11), (12) and (10), respectively.

Irrelevance of the Money Market

We shall now show that the equilibrium conditions given by Eqs. (2), (4) and (6) imply equality of demand for money and supply of money. Substituting Eqs. (6) and (4) into Eq. (2), we get

$$Y = C(Y) + \left(\frac{1-\rho}{\rho}\right)\frac{dH}{P} + \frac{dH}{P} - \bar{K} \Rightarrow$$

$$(Y - C(Y)) + \bar{K} = \frac{dH}{P}$$

The LHS of the above equation constitutes households' saving plus net inflow of foreign capital. It, therefore, represents domestic households' and foreign investors' demand for additional money or additional bank deposit, as they hold their entire saving/investment in the form of bank deposit/money. The RHS gives the supply of additional bank deposit/money. This ensures equality of demand for money and supply of money. Thus, when Eqs. (2), (4) and (6) are satisfied, money demand and money supply become automatically equal.

Evaluation of the Model

This simple model redresses all the major deficiencies of the characterisation of the financial sector in the IS-LM-based open economy macro-models. These models do not show how different kinds of expenditures are financed or how saving generates new credit. Nor do these models consider financial intermediaries, which play a major role in mobilising savings and making them available for financing different kinds of expenditure. The present model incorporates financial intermediaries and brings out clearly the interrelationships that exist among the processes that generate income, saving, new credit and expenditure. It shows that the multiplier process that occurs in the real sector and the money or credit multiplier process that occurs in the financial sector take place simultaneously reinforcing each other. It brings to the fore the process through which savings are used by the financial intermediaries to extend credit.

Unlike the IS-LM-based open economy macro-models, which cannot handle the situation where interest rates are rigid, this model can handle the situation where the interest rates are flexible as well as the one where interest rates are fixed, even though we have not considered the latter case here. The present model can easily be extended to accommodate that case.

Here, we have kept P unchanged. We can easily drop this assumption and explicitly consider the process that determines P. We shall now apply this model to explain the Greek crisis.

3 Greek Crisis

Greece entered into a severe recession since 2008 (see Table 1). We shall use the model developed above to explain this crisis. Along with the severe recession, Greece also found that it was unable to honour its sovereign debt service commitments in 2008. This is another aspect of the Greek Crisis. In what follows, we shall seek to explain both these aspects of the crisis. Obviously, the two are intimately related to one another. In fact, the severe contraction in Greek GDP since 2008 was a major cause of the sovereign debt crisis.

The currency of the Greece economy is euro. It is a currency, which the Greece government cannot print. It is issued by the European Central Bank (ECB). The stock of euro in the possession of Greece constitutes the stock of high-powered money in the possession of Greece. It is held as reserve of commercial banks of Greece and

currency by the non-bank public. We assume for simplicity that there is no currency holding by non-bank public. The non-bank public hold all their savings as bank deposits. We further assume for simplicity that the bank deposits are the only kind of financial asset available in the economy. Domestic households and foreign investors invest in bank deposits only. The banks in Greece can give new loans only if they get more high-powered money. The only source of new high-powered money to Greece is net exports (NX) and net inflow of capital(K). We regard K to be net of the interest payments made by the banks to the foreigners. We assume the net inflow of capital K to be autonomous, and its value is denoted by $\bar{K}$. We assume for simplicity that Greece trades only with other Eurozone countries so that $\bar{e} = 1$. The option of financing expenditure by borrowing from the central bank is also not available to the Greek Government. Therefore, the supply of new high-powered money in Greece is given by Eq. (4), with $\bar{G} = 0$. Thus, the model that we use to explain the Greek crisis is the one that we have developed above, with $\bar{e} = 1$ and $\bar{G} = 0$.

The Performance of the Greek Economy in the Pre-Crisis Era and the Beginning of Crisis in Greece

We shall now use the model developed above to explain the growth performance of the Greek economy in the pre-crisis era and the outbreak of crisis in Greece. From the data presented in Table 1, we find that there took place a decisive break in the growth performance of Greece in 1997. During the period 1981–1996, Greece was an extremely slow growing or almost a stagnant economy. In 1997, the growth rate jumped to a high level and remained at such high levels until 2007 (see Table 1). How do we explain this jump in Greece's growth performance? It is quite easy to explain this. From Table 1 we also find that growth rates in the USA and in many Eurozone and European countries such as France, Germany, Spain, Ireland and UK increased remarkably during the high growth phase of Greece. Capital also began to flow into Greece from the year 2000, and net capital inflow grew at a very high rate during 2003–2007. In terms of our model, Y^* grew at a high rate during the high growth phase of Greece and in the later half of the period along with Y^*, K also grew at a high rate. We shall now examine how an increase in Y^* and $\bar{K}$ affects the growth rate of GDP in our model.

*3.1 Effect of an Increase in Y**

We shall use Fig. 3 to examine how an increase in Y^* affects growth rate. In Fig. 3, initial equilibrium values of Y and r, denoted Y_0 and r_0, correspond to the point of intersection of IS and LL schedules representing Eqs. (2) and (6), respectively. In the lower panel, HH represents Eq. (4) and the initial equilibrium $\frac{dH}{P}$ is denoted by $\left(\frac{dH}{P}\right)_0$. Let us now examine how IS and LL shift following an increase in Y^*. Let us first focus on IS representing Eq. (2). Following an increase in Y^*, net export rises bringing about excess demand for goods and services at every (Y, r) on the initial IS. Hence, corresponding to any given r, the goods market will be in equilibrium at

a larger Y. Hence, the IS shifts to the right. Let us now focus on the LL schedule representing Eq. (6). Following an increase in Y^*, there emerges excess supply of credit at every (Y, r) on LL. Hence, corresponding to any given r, the credit market will be in equilibrium at a larger Y. Accordingly, LL shifts to the right. The new IS and LL are labelled IS_1 and LL_1, respectively. The rightward shift in the IS corresponding to the initial equilibrium (Y, r) will be less than that in LL. Let us explain. Following an increase in Y^*, at the initial equilibrium (Y, r), net export becomes larger bringing about an excess demand for goods and services. Corresponding to any given r, the goods market will now be in equilibrium at a larger Y. However, as follows from Eq. (2), at this larger Y, net export must be larger, since the increase in C is less than that in Y. Again, at the initial equilibrium (Y, r), there emerges excess supply of credit following the increase in net export caused by the rise in Y^*. Hence, as follows from Eq. (6), at the initial equilibrium r, the credit market will be in equilibrium at a larger Y. At this larger Y, as we find from Eq. (6), net export is at its initial equilibrium value. This means that at the initial equilibrium r, the rightward shift in LL is larger than that in the IS. In the new equilibrium, therefore, Y is larger and r is less.

We may describe the adjustment process as follows. Following an increase in Y^*, at the initial equilibrium (Y, r), net export goes up bringing about an excess demand for goods and services and excess supply of credit. The former will raise Y, while the latter will induce a fall in r. Equilibrium in the goods and the credit market will be restored through the increase in Y and the fall in r. This analysis yields the following proposition:

Proposition 1 *A higher growth rate in Y^* brings about a higher growth rate in Y and a larger fall in r.*

3.2 Effect of an Increase in $\bar{K}$

Let us now examine how an increase $\bar{K}$ affects Y and r. This we do with the help of Fig. 4, where the initial equilibrium values of Y and r, labelled Y_0 and r_0, respectively, correspond to the point of intersection of IS and LL representing Eqs. (2) and (6), respectively. Following an increase in $\bar{K}$, as follows from Eqs. (2) and (6), IS remains unaffected, while LL shifts to the right. Thus, Y rises and r falls. These results can be easily derived mathematically. The above analysis yields the following proposition:

Proposition 2 *A higher growth rate in K brings about a higher growth rate in Y and a larger fall in r.*

From Propositions 1 and 2, it follows that the higher growth rate in Greece since 1997 was due to higher growth rates of GDP in other European countries and USA and higher rate of growth of inflow of capital.

It also follows from above that the higher growth rates in other European countries and USA and the higher growth rate in net inflows of capital brought about sharp falls

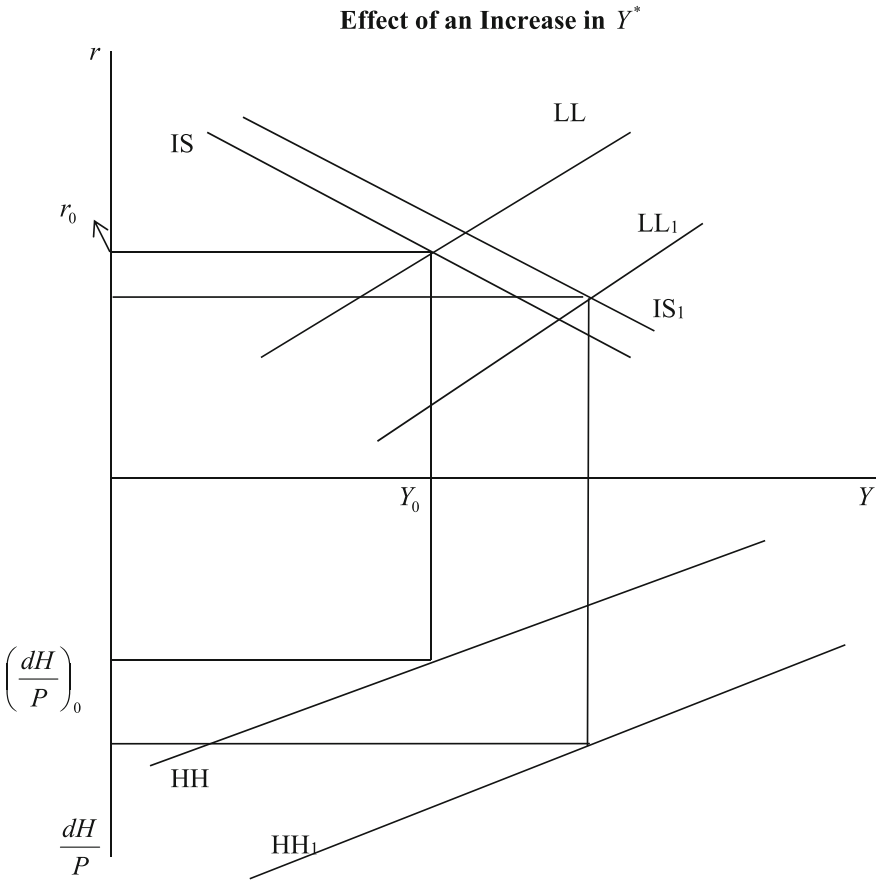


Fig. 3 Effect of an increase in Y^*

in interest rates. This induced the Greek Government to borrow on a large scale to finance additional expenditure. This explains how Greek Government accumulated a sizable amount of debt by the beginning of 2008.

The Crisis in Greece since 2008

The crisis that engulfed Greece since 2008 can also be explained using Propositions 1 and 2. From the data given in Tables 1 and 2, we find that the factors that turned favourable since 1997 became extremely unfavourable since 2008. There took place a sharp decline in the growth rates of GDP in all European countries and the USA following the collapse of real estate bubbles. In fact, GDPs contracted in most countries. Capital also instead of flowing in started flowing out of Greece. Both these factors precipitated the severe recession that Greece slipped into since 2008. Interest rates also shot up steeply. As we have already mentioned, Greek Government accumulated sizable debt by the beginning of 2008. With the large contraction in

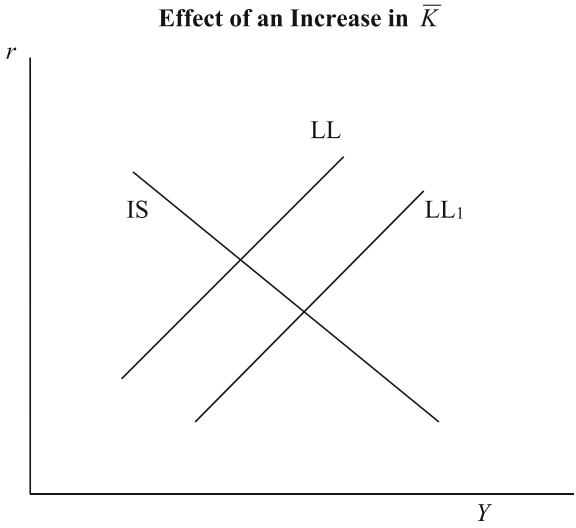


Fig. 4 Effect of an increase in $\bar{K}$

Greek GDP in 2008 (see Table 1), sovereign debt GDP ratio increased steeply in Greece. Foreign investors' not only stopped lending to Greece but also started withdrawing the funds they invested in Greece. With the contraction in GDP, revenue of the Greek Government declined substantially. With the withdrawal of deposits on a large scale, commercial banks in Greece got into serious trouble and they sorely needed recapitalisation. As a result of all these factors, the revenue of the Greek Government fell far short of what was needed to service Greek Government's debt. New loans were not available either. To avoid loan default, Greek Government had to seek the assistance of the IMF, European Commission and the European Central Bank in 2010. They obliged, but imposed stringent austerity measures, which led to further contraction of Greek GDP. The austerity induced contraction aggravated Greece's debt woes instead of alleviating them. However, how austerity contributed to Greece's problems is beyond the scope of the present paper.

4 Conclusion

The paper develops a model to explain the crisis in Greece. The IS-LM model-based open economy macro-models are not applicable to Greece, as they do not show how different kinds of expenditures are financed. Nor do they incorporate financial intermediaries. Bernanke and Blinder (1988) incorporated banks in the IS-LM-based macro-models, but they could not resolve the problem relating to the financing of different kinds of expenditure. The model developed here resolves this problem and shows how net inflows of capital into Greece financed additional private and

Table 2 Portfolio equity, net inflows (BoP, current US $) (Portfolio equity includes net inflows from equity securities other than those recorded as direct investment and including shares, stocks, depository receipts (American or global) and direct purchases of shares in local stock markets by foreign investors. Data are in current US dollars.)

Year	1981	82	83	84	85
Greece	0	0	0	0	0
Year	1986	87	88	89	90
Greece	0	0	0	0	0
Year	1991	92	93	94	95
Greece	0	0	0	0	0
Year	1996	97	98	99	2000
Greece	0	0	0	−2,588,600,000	1,636,800,000
Year	2001	02	03	04	05
Greece	1,829,000,000	1,400,461,074	2,568,346,806	4,290,377,439	6,292,635,261
Year	2006	07	08	09	10
Greece	7,529,407,898	10,865,115,709	−5,259,941,990	763,722,456	−1,459,911,258

Source World Bank

government expenditures. It also shows how processes of generation of income, saving, credit and expenditure are inextricably linked together.

Even though the existing literature identifies a surge in capital inflows into Greece and its subsequent drying up as the sole factor responsible for the Greek crisis, we point to an additional factor, namely remarkable jump in the growth rates in many European countries and the USA riding on the waves of the dotcom and real estate bubbles and the severe recession that they went into following the collapse of the asset price bubbles. In sum, the paper attributes the crisis of Greece to the unbridled speculative activities of the global financial capital.

References

Bernanke, B. S., & Blinder, A. S. (1988). Credit, money and aggregate demand. *The American Economic Review, 78*(2), 435–439.

Dellas, H., & Tavlas, G. S. (2013). The gold standard, the euro and the origins of the Greek sovereign debt crisis. *Cato Journal, 33*(3), 491–520.

Dobb, M. (1966). *Soviet economic development since 1917*. New York: International Publishers.

Gibson, H. D., Palivos, T. & Tavlas, G.S. (2014). Thr crisis in the euro area. Journal of Macroeconomics, *38*(B), 233–460.

Krugman, P. (2013). Revenge of the optimum currency area. NBER Macroeconomic Annual, *27*, 439–448 (2012).

Lane, P. R. (2012). The European sovereign debt crisis. *Journal of Economic Perspectives, 26*(3), 49–68.

Jha, S., Ghosh, A., & Ghosh, C. (2016). Interaction between the real sector and the financial Sector: An alternative to the IS-LM model. *Contemporary Issues and Ideas in Social Sciences, 6*(1), 1–14.

Rakshit, M. (1993). Money, credit and monetary policy. In T. Majumdar (Ed.), *Nature, man and the environment*. Delhi: Oxford University Press.

School Language Policy, Crime and the Minority Underclass

Indraneel Dasgupta

1 Introduction

In practically all societies divided between a well-defined majority ethno-linguistic community and at least one such minority, the language and cultural-cum-behavioural conventions to be adopted in public educational institutions constitute a standard site of political contestation between communities. Broadly, two different positions can be perceived in this context. One may term the first position behavioural-expressive centralization or *unitarianism*. This involves the idea that public institutions in general, and public educational institutions in particular, should solely reflect the linguistic and cultural-behavioural conventions of the majority, so that minority individuals may access these institutions only if they 'assimilate', i.e., adopt these conventions. The second position may be termed behavioural-expressive *federalism*. This enjoins public institutions to partially adopt the cultural-linguistic and behavioural conventions of at least the larger minorities, so that the latter may access these facilities while maintaining their linguistic and cultural/behavioural distinctiveness. The objective of this paper is to provide an analytical framework within which these policy stances can be assessed, and their implications for income distribution, decentralized crime and welfare dependency explicated.

The formal model presented in this chapter draws heavily from sections of a much broader analysis carried out in my unpublished discussion paper titled 'Assimilation, criminality and ethnic conflict', co-authored with Diganta Mukherjee of the Indian Statistical Institute (IZA Discussion Paper No. 7924, January 2014). The concrete application to school language policy pursued here was however not attempted in that paper. The primary analytical focus of that paper was on ethnic conflict: an issue I do not engage with at all here.

I. Dasgupta (✉)
Economic Research Unit, Indian Statistical Institute, 203 Barrackpore Trunk Road, Kolkata 700108, West Bengal, India
e-mail: indraneel@isical.ac.in

P. Ray et al. (eds.), *Economics, Management and Sustainability*,
https://doi.org/10.1007/978-981-13-1894-8_6

Behavioural-expressive unitarianism (unitarianism for brevity) may take the extreme form of inserting minorities into a majority institutional setting without giving them the legal scope to opt-out. Legislated removal of minority children on an extensive scale from their parents and communities, and relocation in institutional and foster-care settings involving immersion in the majority language and culture, provides a stark example. In Australia, children of Aboriginal and Torres Strait Islander descent used to be removed from their families by government agencies and church missions, to be brought up in white institutional and foster care. In 1997, following a national inquiry, the Australian Human Rights and Equal Opportunity Commission concluded that between one in three and one in ten indigenous children were forcibly removed during 1910–1970.[1] In Canada, a network of residential schools for children from First Nations, Métis, and Inuit communities was set up with funding from the government's Department of Indian Affairs and administered by churches. The system was primarily active from 1876 until the mid-twentieth century. School attendance was made compulsory and, in some parts, residential schools were the only option. In 2008, public apologies were issued by the Prime Ministers of both countries in their respective Parliaments for past adoption of these policies.

While these cases of physical removal and absorption are extreme, Australia and Canada are not unique in having enforced policies of cultural and linguistic assimilation that are binding on minorities. Perhaps even more pervasive, however, are policies to incentivise individual members of minority communities to embrace majority norms. Language, syllabus and cultural policies followed in public educational institutions, the official language followed in law courts and public administration, language and cultural content of citizenship tests, etc., are all instruments that can and indeed are used to nudge minority individuals towards extensive adoption of majority ethno-linguistic norms, by increasing the relative benefits from doing so.[2] In these cases, minority individuals are notionally free to opt out of assimilation, but only by losing access to valuable public services.

Standard historical examples of behavioural-expressive federalism (federalism for brevity) come from the Austro-Hungarian empire, Yugoslavia and the Soviet Union. In these cases, in regions of minority concentration, many, or even all, public educational institutions adopted the minority's language as the medium of instruction, and at least a large part of both judicial and administrative business was carried out in the minority's language. India, Francophone Canada, and, to a lesser extent, many parts of the US with large Hispanic populations, provide contemporary examples.

What are the relative economic merits of unitarianism vis-à-vis federalism, in our linguistic-cultural context? The former imposes costs of access to public educational institutions on minorities, but makes such access costless for the majority. The latter, in effect, divides up public institutions in general, and the public edu-

[1] *Bringing Them Home – Report of the National Inquiry into the Separation of Aboriginal and Torres Strait Islander Children from Their Families* (Canberra 1997).

[2] Denial of recognition to the Kurdish language in Turkey is linked to the Turkish nationalist policy of cultural assimilation. In Latvia, despite about 40% of the population being Russian-speaking, Latvian remains both the sole state language and a requirement for citizenship. In the UK, English language requirements for citizenship tests have been progressively tightened in recent years.

cation system in particular, between communities, imposing costs on the majority for accessing the minorities' share of public institutions, and vice versa. Cultural-linguistic segregation in the public education system is then largely replicated by the consequent cultural-linguistic segmentation of the labour market. Seen in this light, an important argument in support of the case for unitarianism, and its ultimate objective of assimilating linguistic-cultural minorities to the majority's linguistic and behavioural conventions, appears to be the following. Cultural-cum-linguistic segregation, by leading to socio-economic exclusion, generates a poverty-stricken minority underclass, which puts pressure on the welfare system and/or law enforcement, thereby negatively impacting the majority. Xenophobic political parties often seek to magnify and exploit majority anxieties by simultaneously charging minorities' with both an unwillingness to assimilate and an excessive propensity to engage in crime, and explain away their poverty and exclusion in such terms.[3]

Despite its policy importance, comparative assessment of the impacts of linguistic-cultural centralization and linguistic-cultural federalism, on income distribution, decentralized crime and welfare dependency, has received little analytical attention in the formal theoretical literature on political economics. This paper seeks to address this lacuna.[4]

I consider a society consisting of a majority and a minority. These communities *antagonistically* differ in terms of a set of behavioural-expressive traits and conventions, which are relevant for learning interaction and coordination. Individuals born into a community acquire that community's traits and conventions as part of their upbringing within the community. Language, including dialect, idiom, accent and modes of expression, constitutes the most transparent example of such learning-relevant conventions, but not the only one. Working according to a particular time allocation routine and holiday schedule (e.g. not working on Fridays or Sundays), dress codes (e.g. Islamic veils or Sikh turbans), dietary restrictions (e.g. injunctions against beef, pork, alcohol and non-kosher meat, or adoption of vegetarianism), rules of social interaction (e.g. untouchability or gender-segregation): all constitute common examples of community-specific behavioural traits and conventions that are relevant for within-school interaction and teaching-learning coordination. Expanding on Akerlof and Kranton (2000), I assume that 'switching identity', or bringing one's behaviour and modes of expression into alignment with those commonly present in (and thereby constitutive of) the other community, is feasible but

[3]Contemporary examples include political parties such as the French National Front, the Dutch Party for Freedom, the Bharatiya Janata Party of India, Jobbik of Hungary and Golden Dawn of Greece.

[4]The theoretical contributions most closely related to the concerns of this paper are by Ortega and Tangerås (2008), who develop a political-economic analysis of the imposition of mono-lingual education by dominant groups, and Dasgupta (2017), who examines the connections between linguistic assimilation and group conflict over identity goods. Neither of these two contributions addresses the implications of language and cultural policy in schools for either decentralized crime or welfare dependency. More distantly related are contributions by Lazear (1999), Kónya (2005), Kuran and Sandholm (2008), Li (2013) and Bowles et al. (2014), who develop models of assimilation (or, more generally, social segregation and integration) and that by Akerlof and Kranton (2000), who explain forms of dysfunctional individual behaviour in terms of stresses generated by identity norms.

costly in terms of effort. Individuals vary in terms of their identity switching costs, and are endowed with one unit of effort that they can allocate between learning and identity-switching. There exists one unit of a (non-rival but excludable) pedagogic public good ('schools'), which is allocated across communities according to public policy.

I first consider two alternative policy scenarios under the assumption of secure property rights over income from productive activities. Under the first, unitary, scenario, the entire pedagogic good is costlessly available to all majority individuals, the intuitive interpretation being that the educational system is entirely organised according to the linguistic and behavioural conventions of the majority. To access the educational system, therefore, every minority individual has to incur her idiosyncratic identity switching cost. Individuals' earnings consequent on accessing the educational system are simply unity minus their identity switching cost, the difference being assumed positive. Individuals earn nothing if they don't access the pedagogic public good, the intuitive interpretation being that individuals can only produce after acquiring knowledge and training through the school system.

Under the second, federal, scenario, the pedagogic public good is divided between the two communities according to their population shares. Her own community's share of the public good is costlessly available to a community member, but she cannot access the share of the other community. Intuitively, this formulation is meant to capture a situation where some parts of the educational system (say, schools located in areas of minority concentration) use the minority's language as the medium of instruction and adopt its behavioural conventions (say, gender-segregated and burkha-permitting class rooms, halaal meat-only cafetarias, holidays on Fridays rather than Sundays and religious instruction as part of the curriculum), while the rest use the majority's language and behavioural conventions (which violate those of the minority). A community's share of the school system is identical to its population share. An individual's earnings are simply given by the size of the segment of the public good allocated to her birth community. The interpretation is that, due to the presence of administrative indivisibilities, being able to access a larger segment of the education system implies being able to choose from larger menus of pedagogic styles, disciplinary combinations or specializations, infrastructural facilities and school locations, which facilitates a closer match between an individual's idiosyncratic learning aptitude or intrinsic comparative advantage and the training or education she acquires. Receiving an education more appropriate to one's idiosyncratic characteristics implies better learning outcomes, which in turn generate higher personal productivity.

Thus, in sum, a unitary school system implies a better fit between a minority individual's personal learning-relevant characteristics and the education she receives, compared to a federal one. This improves her productivity. This positive effect is counteracted by the identity-switching effort cost she has to incur, which reduces her productivity. I assume that the net effect on productivity is positive for some, but not all, minority individuals, and positive for the minority community on average.

I first identify a set of parametric restrictions under which a federalist education system, once installed by fiat or as a consequence of political contestation between

communities, can be self-sustaining, in that no individual will have an incentive to unilaterally shift to the other community's segment of the system (i.e., switch identity). Nonetheless, under these parametric conditions, both communities would achieve aggregate income gains if the state were to shift to a unitary educational system organised according to the majority's linguistic and behavioural conventions by fiat. Thus, under my parametric restrictions, enforced assimilation by the minority to the majority's behavioural-expressive norms worsens the income distribution within the minority community, even as it makes both communities better off on average. In this sense, my benchmark model formalizes and clarifies the efficiency case for a unitary school system, while also highlighting its adverse equity consequences.

I proceed to examine how my efficiency conclusions hold up under imperfect property rights protection. I extend the analysis by incorporating individual expropriation as a way of acquiring income, in addition to school-mediated production. I conceptualize expropriation primarily as competitively determined returns from unproductive criminal activities ('theft'), but possibly including legally enforced social transfers (welfare benefits) to non-productive individuals, funded by taxes on productive ones. I show that a unitary education system generates, as a stable equilibrium phenomenon, an unproductive underclass dependent on expropriation. This underclass exhibits a disproportionately high presence of the minority. Assimilation may both immiserize and criminalize *every* member of the minority community, even when, *sans* expropriation, it would both generate income gains for that community on average and make a large proportion of its members better off. Expropriation may however be entirely absent under a federal education system. Thus, the aggregate income gain for the minority community brought about by linguistic-cultural centralization may be more than fully negated by the decentralized distributive conflict it generates, via its dis-equalizing impact on income distribution within that community. The extent of such negation depends on how strongly property rights are protected: therefore, reducing social losses due to expropriation requires greater spending on prevention of property crimes. Under a federal education system, however, even weak property rights protection may suffice to eliminate expropriation. Hence, the productivity case for assimilation needs to be qualified by its causal connection with distributive conflict and the creation of an unproductive minority underclass, while the equity case remains dubious. I thus provide a priori grounds for adopting a cautionary position with regard to integrationist policy claims.

Section 2 sets up the benchmark model, under the assumption of secure property rights over income from productive activities. Section 3 incorporates expropriation as an alternative avenue of income generation. Section 4 concludes.

2 The Model with Secure Property Rights[5]

Consider a population of size normalized to 1, comprised of two groups, M (majority) and N (minority), with population shares m and n respectively, $m = (1 - n)$, $\mathrm{n} \in \left(0, \frac{1}{2}\right)$. Each member of the population is endowed with one unit of effort, which she expends on activities related to earning income.

To earn income, each individual needs to acquire education via a school system. In order to access the school system, each individual needs to acquire some identity-related, or community-specific, linguistic-cultural characteristics, to successfully engage in learning-related ('class-room') negotiations and coordination. The marginal product of effort, contingent on acquiring the characteristics specific to community $i \in \{M, N\}$, is θ_i, where $\theta_i \in [0, 1]$ is the proportion of a composite unit of educational institutions that is organised according to the cultural-cum-behavioural conventions and characteristics of community i. Thus, the benefit from acquiring a particular set of expressive and behavioural conventions depends positively on how pervasive those conventions are in educational institutions. This captures the idea that being able to access a larger segment of the school system implies better learning outcomes, which in turn generate higher personal productivity. Given any community $i \in \{M, N\}$, I shall denote the other community by $-i$. For j born into community i, acquisition of the behavioural and expressive conventions of her own community is costless (reflecting socialization in childhood), but acquisition of those of the other community involves an 'identity switching' cost, modelled as an effort cost c; c is idiosyncratic and distributed over $[\rho_i, \bar{\rho}_i]$, with $0 < \rho_i < \bar{\rho}_i < 1$, according to some continuous and differentiable distribution function $F^i(c)$.

An obvious interpretation of c is in terms of the effort spent in acquiring a new language and behavioural rules instead of substantive knowledge, techniques, and modes of problem-solving within a specific discipline: some are inherently more efficient learners of language and 'manners'. A deeper one is that not all can internalize alien conventions equally. The degree of functionality within the context of a set of culturally/linguistically alien rules varies across persons born into the same community, leading to idiosyncratic differences in learning outcomes and consequently market productivity. These differences are however not intrinsic but specific to the cultural construction of educational institutions: these differences would disappear if production-enabling educational institutions were organized according to the conventions one was originally socialized into.[6] In any case, the formal upshot is that, for j born into community $-i$, the return from adopting the learning relevant behavioural conventions of the other community, i, is $\theta_i\left(1 - c_{-i,j}\right)$, where $c_{-i,j}$ is

[5]The benchmark model developed in this section is broadly similar to that presented by Dasgupta (2017), though the substantive questions investigated there are very different.

[6]This is a familiar general idea in the sociology of education, exemplified by the various writings of Pierre Bourdieu. See, for example, Bourdieu and Passeron (1977). The Chilean film Machuca (2004), written and directed by Andrés Wood, provides a striking portrayal of identity switching costs imposed on poor Native American children when they are enrolled on scholarship in an exclusive private school with an almost entirely White and upper class student body, in the context of Chile in 1973.

the identity-switching (marginal) effort cost of negotiating an alien educational environment for the individual.[7] For such an individual, the return from persisting with one's original behavioural conventions is $(1 - \theta_i)$. I assume that the distribution of identity switching costs follows an exponential form:

$$F^i(c) = (\bar{\rho}_i - \rho_i)^{-\alpha_i}(c - \rho_i)^{\alpha_i}; \tag{1}$$

where $\alpha_i > 0 \forall i \in \{M, N\}$. In case of a concave cost distribution ($\alpha_i \in (0, 1)$) more than half the minority population falls below the mid-point of the cost distribution. Thus, intuitively, minority individuals are more likely to be low cost, rather than high cost; or, equivalently, concentrated in the lower part of the cost distribution with regard to assimilation. The opposite holds for a convex cost distribution. Thus, a concave cost distribution would appear, a priori, to be the case where assimilation is most likely to benefit the minority community on average. Contingent on switching identity, the income $I_{-i,j}$ of j born into community $-i$ falls in the interval $\left[\theta_i(1 - \bar{\rho}_{-i}), \theta_i(1 - \rho_{-i})\right]$.

Let n_M be the size of the 'assimilated' minority population (those who adopt the behavioural and expressive conventions of the majority despite being brought up in the minority community); $n_M \in [0, n]$. Then the assimilation cost of the marginal assimilated member of N is given by:

$$\breve{c}(n_M) \equiv F^{N^{-1}}\left(\frac{n_M}{n}\right). \tag{2}$$

$\breve{c}(.)$ is the inverse supply function for assimilated individuals: if the population size of N individuals who rationally assimilate is n_M, then the highest cost incurred must be exactly $\breve{c}(n_M)$. By (1) and (2):

$$\breve{c}(n_M) \equiv \left(\frac{n_M}{n}\right)^{1/\alpha_N}(\bar{\rho}_N - \rho_N) + \rho_N; \tag{3}$$

so that

[7]Generalized discrimination against the minority can be modelled as a constant cost component, $d \leq \rho_N$, that impacts all assimilating N individuals equally. Thus, an increase in such discrimination simply reduces the returns from assimilation by an identical amount ($\theta_1 d$) for all minority individuals. Individuals may perceive their own expressive and behavioural habits as norms rather than conventions, in that they may intrinsically value them as ideals to live by. In that case, identity-switching will involve a psychic cost. If such marginal psychic cost increases with the level of workplace effort, individuals may rationally provide less than full effort in an alien work environment. The effort level provided will then vary according to idiosyncratic differences in the marginal psychic cost function. Though evidently compatible with my analysis, I refrain from explicitly modelling this additional source of idiosyncratic differences in productivity on considerations of expositional ease and simplicity.

$$\breve{c}'(n_M) = \frac{(\bar{\rho}_N - \rho_N)}{n\alpha_N}\left(\frac{n_M}{n}\right)^{\frac{1-\alpha_N}{\alpha_N}} > 0 \text{ for all } n_M \in (0, n]; \tag{4}$$

$$\breve{c}''(n_M) = (1-\alpha_N)\frac{(\bar{\rho}_N - \rho_N)}{(n\alpha_N)^2}\left(\frac{n_M}{n}\right)^{\frac{1-2\alpha_N}{\alpha_N}}. \tag{5}$$

Thus, the marginal assimilation cost function (or the inverse supply function) $\breve{c}(.)$ is *increasing* in the size of the assimilated population over $(0, n]$. It is convex if $\alpha_N \in (0, 1)$ and concave if $\alpha_N > 1$. Analogous expressions hold for M.

Assumption 1 (i) $[\bar{\rho}_N > m > \rho_N > m - n]$; and (ii) $[m > E(c_N)]$.

By Assumption 1(i), when the educational public good is divided according to population proportion, no individual will unilaterally choose the other community's expressive and behavioural conventions. Thus, Assumption 1(i) ensures that, if a population-proportionate federal school system was brought about by state fiat, it would be self-perpetuating, since no individual would have a unilateral incentive to migrate to the other community's education sector. Such a system would constitute a Nash equilibrium. Hence, a shift to a unitary system would require a purposive act of policy intervention. Assumption 1(i) also embeds the analytically more interesting and empirically more plausible scenario where some, but not all, minority individuals would be better off if the entire educational system was organised according to the majority's conventions. Since I wish to highlight the role played by property rights protection in determining the social consequences of a unitary education policy, I set up the most favourable scenario for a unitary education system under secure property rights by assuming that the aggregate benefit to the minority from linguistically-culturally unifying the school system along majority conventions is greater than its total cost (Assumption 1(ii)).

What happens if, from an initial condition of population-proportionate federalism constituting a stable individually-rational equilibrium, public policy shifts the school system to a unitary form organised according to the majority's linguistic-cultural conventions, under secure property rights? I now turn to this question. I ignore the possible case where the unitary school system is organised according to the minority's conventions since the treatment is symmetric. The answers are summarized in Observation 1 below, which follows immediately from Assumption 1.

Observation 1 *Let Assumption 1 hold. Then, under linguistic-cultural unitarianism according to the majority's conventions relative to linguistic-cultural federalism:*

(a) *every member of the majority community earns more,*
(b) *total income in society is higher,*
(c) *total income of the minority community rises, and*
(d) *some, but not all minority individuals suffer an absolute income reduction.*

Observation 1 articulates the efficiency argument for linguistic unitarianism under a best-case scenario. Every member of M gains income if N assimilates. The economies of scale assimilation generates outweigh the costs of integration incurred by the latter, so that total income of society increases. However, while incomes within a community are identical under segregation, reflecting equal inherent productivity, idiosyncratic differences in the ability to function within an alien culture opens up income inequality inside N when it assimilates (though incomes within M remain equalized). Nonetheless, N benefits monetarily on average from assimilation since the gain from assimilation is greater than the average cost. The larger the majority relative to the minority's average cost of assimilation, the higher the gain to minority individuals on average from assimilation. However, since the upper bound on assimilation costs is higher than the gain from assimilation, a positive proportion of N individuals (those with costs in $(m, \bar{\rho}_N)$) must suffer a fall in income under assimilation, while those with lower costs, i.e. costs in the interval (ρ_N, m) will achieve income gains.

3 School Policy Under Insecure Property Rights

From the perspective of the minority, the key justification for assimilation identified by my analysis so far is its positive impact on the earnings of those minority individuals whose identity adjustment costs are low relative to the gain from assimilation. I now proceed to show that these purported gains may be illusory: they may be more than eliminated by decentralized conflict over expropriation generated endogenously by assimilation when property rights over income from production are insecure.

I interpret expropriation primarily as illegal income from individual participation in a competitive criminal sector that involves extortion, theft and robbery. More broadly, however, it may involve legislated redistributive transfers (welfare payments) to non-producers as well. I model expropriation as a lump-sum tax on all producers: the size of this tax rises with the relative size of the population engaged in expropriation, till some ceiling. The expropriation sector is competitive, in that entry is free, all expropriators act as price-takers and earn identical returns from expropriation.

Expropriation yields $r, r = R$ if the proportion of the population engaged in it, x, is not more than $x^* \in (0, 1)$. The most that a producer can lose to expropriators is $\bar{L} \in (0, n)$. Both R and $\bar{L}$ are to be thought of as measures of property rights protection. For crime, I interpret R as the most that an individual criminal can extort, and $\bar{L}$ as the amount a producer cannot defend, given the policing and legal structure. The former binds when the criminal population is sufficiently small (below x^*). The latter binds at x^* and beyond. Expansions in the criminal population beyond x^* accordingly reduce earnings in that sector.[8] When expropriation involves welfare transfers to non-producers, R represents the most that a given political system can

[8]This formulation is similar to that of Murphy et al. (1993). However, they do not address identity aspects at all, which constitute our explicit focus. This leads to a substantive difference in consequences. While absence of expropriation constitutes a locally stable equilibrium in their model, the

provide. If the claimant population is small, the system accommodates additional claimants by increasing the tax rate, rather than by reducing per capita benefits. Once the tax ceiling is reached, further increases in the population of transfer claimants lead to a commensurate reduction in per capita benefits. Thus, for a productive individual, loss from expropriation is $L = Min\{\frac{xR}{(1-x)}, \bar{L}\}$, while individual gain from expropriation is given by:

$$\begin{aligned} r &= R \, if \, x \leq x^* \equiv \bar{L}/(R + \bar{L}); \\ &= \frac{(1-x)\bar{L}}{x} \quad \text{otherwise.} \end{aligned} \tag{6}$$

Given any proportion of the population engaged in production $(1 - x)$, let $\pi_P(x)$ be the *minimum* net income possible such that there exists a set of individuals with measure x, all members of which earn $\pi_P(x)$ or less *in excess of* r from production. Under linguistic unitarianism, recalling (2),

$$\begin{aligned} \pi_P(x) &= \left[1 - Min\left\{\frac{xR}{1-x}, \bar{L}\right\} - \breve{c}_N(n-x)\right] - r \text{ if } x \leq n; \\ &= \left[1 - Min\left\{\frac{xR}{1-x}, \bar{L}\right\}\right] - r \text{ if } x > n, \end{aligned} \tag{7}$$

whereas, under linguistic federalism,

$$\begin{aligned} \pi_P(x) &= \left[n - Min\left\{\frac{xR}{1-x}, \bar{L}\right\}\right] - r \text{ if } x \leq n; \\ &= \left[m - Min\left\{\frac{xR}{1-x}, \bar{L}\right\}\right] - r \text{ if } x > n. \end{aligned} \tag{8}$$

Analogously, let $\bar{\pi}_P(x)$ be the *maximum* net income possible such that there exists a set of individuals with measure $(1 - x)$, all members of which earn $\bar{\pi}_P(x)$ or more *in excess of* r from production. Evidently, $\pi_P(x) = \bar{\pi}_P(x)$ if [either $x < n$ or $x > n$], while $\pi_P(n) < \bar{\pi}_P(n)$. A level of expropriation x_E is an equilibrium iff $[\pi_P(x_E) \leq 0$ and $\bar{\pi}_P(x_E) \geq 0]$. An equilibrium x_E is (locally) stable iff for some $\varepsilon > 0$, $[\pi_P(x) > 0$ whenever $x \in (x_E, x_E + \varepsilon)$, and $\pi_P(x) < 0$ whenever $x \in (x_E - \varepsilon, x_E)]$.

incorporation of identity switching costs rules out this possibility when assimilation occurs in my model (see Proposition 1(b) below).

Proposition 1 *Let Assumption 1 hold, and let* $[1 - \bar{\rho}_N < R < n]$. *Then:*

(a) under linguistic federalism, absence of expropriation constitutes a locally stable equilibrium; but

(b) under linguistic unitarianism, absence of expropriation cannot constitute an equilibrium, and the minority community must participate proportionately more in expropriation than the majority community in any equilibrium; furthermore, at least one (locally) stable equilibrium involving expropriation will necessarily exist.

Proof of Proposition 1 (a) Since assimilation costs are 0 under federalism, part (a) of Proposition 1 is self-evident.

(b) Suppose under unitarianism no exprop1qqriation is an equilibrium. Then the proportion of N earning at least R is unity. But, as $R \in (1 - \bar{\rho}_N, 1)$, this cannot be. Now, if the entire population expropriates, then the return to it is 0, while the return to production, $1 - \bar{L}$, is positive. Hence (recalling that expropriation must obtain), in any equilibrium, both production and expropriation must engage positive proportions of the population. Evidently, if any M individual is better off through expropriation, then the same must hold for *all* N individuals. Thus, any equilibrium where a positive proportion of M expropriates must also be one where all of N expropriates. Hence, any equilibrium must fall in one of exactly two categories: (a) only N individuals expropriate, or (b) all of N, and some, but not all, of M expropriate. Hence N participates proportionately more in expropriation.

I now show that there exists at least one locally stable equilibrium under a unitary school policy. By (7), $\pi_P(0) = (1 - \bar{\rho}_N) - R < 0$;$\pi_P(1) = \left(1 - \bar{L}\right) > 0$; $\pi_P(x)$ is continuous and identical to $\bar{\pi}_P(x)$ in $[0, n)$ and $(n, 1]$, though discontinuous at $x = n$. Then a stable equilibrium between 0 and n must exist if $\pi_P(n) > 0$, while one lying between n and 1 must exist if $\bar{\pi}_P(n) < 0$. If $[\pi_P(n) \leq 0$ and $\bar{\pi}_P(n) \geq 0]$, $x_E = n$ must be an equilibrium. If $[\pi_P(n) < 0$ and $\bar{\pi}_P(n) > 0]$ then, by continuity of both in $[0, n)$ and $(n, 1]$, $x_E = n$ must be stable. If $\pi_P(n) = 0$, then $x_E = n$ is stable when there exists $\varepsilon > 0$ such that $\pi_P(x) < 0$ for all $x \in (n - \varepsilon, n)$. If there exists $\varepsilon > 0$ such that $\pi_P(x) > 0$ for all $x \in (n - \varepsilon, n)$, then, by continuity, there must be a stable equilibrium $x_E \in (0, n)$. Again, by continuity, the only remaining possibility is that, for some $\varepsilon > 0$, $[\pi_P(x) = 0$ for all $x \in (n - \varepsilon, n)]$. It is easy to check from (7) that this cannot be. Hence, there must exist at least one locally stable equilibrium $x_E \in (0, n]$ whenever $[\pi_P(n) = 0$ and $\bar{\pi}_P(n) > 0]$. By an exactly analogous argument, there must exist at least one locally stable equilibrium $x_E \in [n, 1)$ whenever $\bar{\pi}_P(n) = 0$

■

By Proposition 1(a), universal individual acceptance of the extant distribution of income can co-exist with linguistic-cultural segregation in the education system, as a locally stable equilibrium, when the maximum possible returns from expropriation are low, relative to the size of the minority. Thus, when a minority is relatively populous, and property rights are well protected, dependence on criminal activities and/or welfare transfers may be negligible when the communities are segregated at

the school level. This will also constitute the only possible equilibrium when property rights are sufficiently well protected, so that $R < \left(n - \bar{L}\right)$. Thus, private incentives suffice to eliminate individualized distributive strife over material resources altogether, even though the society can offer only imperfect protection to the property rights of producers. Indeed, even property rights protection that appears minimally effective to N producers, in the sense of providing only an arbitrarily small margin over the return from expropriation, suffices to ensure a locally stable equilibrium that eliminates decentralized distributive conflict under linguistic federalism (n may exceed R by an arbitrarily small amount).

In contrast, under linguistic unitarianism or centralization, even if property rights are 'almost perfectly' protected (R is less than what all but an arbitrarily small proportion of minority individuals can earn from production), it is impossible to eliminate expropriation as an equilibrium outcome (Proposition 1(b)). Due to identity switching costs, linguistic centralization creates an 'underclass' of minority individuals: the proportion of the minority population with earnings arbitrarily close to $1 - \bar{\rho}_N$ is always positive. Hence, some N individuals always find it rational to expropriate. This however reduces the return from production, inducing even more individuals to expropriate. Thus, even a highly effective system of property rights protection does not guarantee that distributive tensions will be negligible: a low value of R is compatible with high levels of expropriation in every equilibrium involving assimilation by all minority producers. In sum, identity switching costs can magnify even minor breaches of property rights protection into high and persistent levels of distributive strife.

Proposition 1(b) also suggests that identity costs create a disproportionately low presence of N in production. Every equilibrium exhibits a relatively high engagement of N in expropriation: thus, the underclass, i.e. those surviving on criminal earnings or welfare transfers, must disproportionately include N individuals. Indeed, in equilibrium, the entire N community may expropriate while the entire M community produces. Paradoxically, despite being the expropriators, all N individuals may suffer income losses on assimilation. Conversely, despite being the expropriated, all M individuals may achieve income gains. Thus, assimilation may causally generate both widespread *immiserization and criminalization* within the N community; indeed this may occur even when potential income gains from assimilation are sizeable for the minority.[9] The following example illustrates this point.

[9] At a broad interpretative level, this finding serves to make sense of the case of the so-called 'criminal tribes' in colonial India. In 1871, the British colonial authorities in India enacted the Criminal Tribes Act, under which communities were defined as habitually criminal and systematically registered. Restrictions on movements were imposed and adult male members were forced to report weekly to the local police. At Independence in 1947, 13 million people in 127 communities faced constant surveillance, mandatory fingerprinting, search and arrest without warrant if found outside prescribed areas. The Act was repealed in 1949. The Act essentially covered marginal communities of itinerant petty traders, pastoralists, gypsies, and hill and forest dwelling tribes, whose life-styles and cultural habits did not conform to the model of settled agriculture, waged labour and commercial exploitation of forest resources that the colonial state was promoting. It was thus an attempt to forcibly assimilate these marginal communities into the state's preferred mode of socio-economic organization. Accordingly, special 'settlements' were constructed for these communities, and many

Example 1 Let $\bar{\rho}_M = \bar{\rho}_N = 0.71, \rho_M = \rho_N = 0.59, m = 0.7, \alpha = 1, R = 0.295, x^* = \frac{1}{2}$. Then Assumption 1 holds, $R = \bar{L}$, $[n > R > (1 - \rho_N) - \frac{nR}{(1-n)}]$, $[1 - \bar{L} > R]$ and $[m < 1 - \bar{L}]$. Given a unitary school system, a stable equilibrium exists where all N individuals expropriate while all M produce. All M individuals earn $(1 - \frac{nR}{(1-n)})$, which is more than m; but all N earn R, which is less than n. However, since $n > R$, no expropriation constitutes a stable equilibrium under a population-proportionate federal school system. Expropriation thus leads to *all* N individuals earning less under linguistic-cultural unification of the education system than what they may have done under a segregated one, though all M earn more. *Sans* expropriation, unification would have generated income gains for approximately 91.7% of the N population, and also increased its total income.

4 Discussion and Concluding Remarks

This paper has developed a theoretical framework within which one may examine the case for linguistic-cultural unification of the educational system in societies with multiple ethno-linguistic communities. I have shown that possible efficiency gains from unification have to be balanced against the consequences of integration expanding income inequality within the minority community. Such expansion may set in motion attempts to expropriate productive individuals which, through cumulative causation, may more than dissipate any income gains accruing to the minority community from integration. Thus, the efficiency case for a unitary education policy needs to be qualified by the possibility of both immiserization and criminalization of the minority. Furthermore, measures to protect property rights, which are resource consuming, may be more relaxed, and hence less costly, under a school system organised on principles of linguistic-cultural federalism, without necessarily generating crime or distributive conflict. Such costs offer an additional caveat against enforced assimilation, and provides conditional support for a federal school policy in the presence of *large* linguistic minorities. However, my analysis also shows that, for relatively small minorities, educational segregation can causally generate high levels of poverty and criminalization, both of which may be reduced by cultural-cum-linguistic assimilation.[10] Nonetheless, assimilation may be blocked by the segment within the minority which loses out from assimilation: the minority community may

were settled (i.e. confined) in villages under police guard, whose job was to ensure that no registered member of the tribe was absent without notice. The Amendment of 1897 empowered local governments to establish separate 'reformatory' settlements, for tribal boys from age four to eighteen, away from their parents (as in Canada and Australia, see Sect. 1). The usually desperate living conditions in these settlements forced significant sections of these communities to take to petty theft and robbery as a means of survival, which reinforced discrimination and exclusion from productive activities brought about by the Act. A vicious cycle of immiserization and criminalization was thereby created, the effects of which persist even now. See Radhakrishnan (2001).

[10]Formally, this is the case where $R > n$ (recall Proposition 1).

itself get split between those who wish to assimilate and those who do not.[11] Thus, a small marginalized minority may end up in a culturally and linguistically ghettoised 'identity trap' associated with high levels of crime, poverty and low intensity but persistent internal conflict. Even if integrationist efforts are successful, a society may end up with a permanent underclass comprising disproportionately of individuals from minority origins, surviving precariously through various combinations of petty criminality and welfare dependency, simultaneously as other minority individuals integrate and achieve income gains.[12] Elsewhere (Dasgupta 2017) I have discussed in detail how various language policy measures may be envisaged to reduce identity switching costs for minority individuals. Detailed formal examination of such policy measures would appear to constitute a useful avenue of future research, especially in the context of linguistic-cultural identity traps.

References

Akerlof, G. A., & Kranton, R. E. (2000). Economics and identity. *Quarterly Journal of Economics, 115*(3), 715–753.

Austen-Smith, D., & Fryer, R. G., Jr. (2005). An economic analysis of 'acting White'. *Quarterly Journal of Economics, 120*(2), 551–583.

Bourdieu, P., & Passeron, J.-C. (1977). *Reproduction in education, society and culture, translated by Richard Nice*. London: Sage.

Bowles, S., Sethi, R., & Loury, G. (2014). Group inequality. *Journal of the European Economic Association, 12*(1), 129–152.

Dasgupta, I. (2017). Linguistic assimilation and ethno-religious conflict. In W. Buchholtz & D. Ruebbelke (Eds.), *The Theory of Externalities and Public Goods: Essays in Memory of Richard C. Cornes* (pp. 219–242) Berlin: Springer.

Konya, I. (2005). Minorities and majorities: a dynamic model of assimilation. *Canadian Journal of Economics, 38*(4), 1431–1452.

Kuran, T., & Sandholm, W. (2008). Cultural integration and its discontents. *Review of Economic Studies, 75,* 201–228.

Lazear, E. (1999). Culture and language. *Journal of Political Economy, 107*(6), S95–S126.

Li, Q. (2013). Language and urban labor market segmentation: Theory and evidence. *Journal of Urban Economics, 74,* 27–46.

Murphy, K., Schleifer, A., & Vishny, R. (1993). Why is rent-seeking so costly to growth? *American Economic Review, 83*(2), 409–414.

[11] Conflicts within the African-American community over 'acting White' constitute a specific example, of which Austen-Smith and Fryer (2005) provide a formalization.

[12] Urban riots in the UK, France and Sweden are all recent reminders of the volatility of this underclass. Conversely, partition of a country along ethno-linguistic lines usually leads to large-scale but *incomplete* ethnic cleansing, leaving behind small minority enclaves which tend to get stuck in the kind of identity traps that I have highlighted. Discrimination by the majority, with or without official sanction, makes these identity traps even harder to escape. This seems to be the case for the Muslim minority in some parts of India, for the Arab minority in Israel, as well as for various local minorities in parts of the former Yugoslavia. The 1989 Hindi film "Salim Langde Pe Mat Ro", directed by Saeed Akhtar Mirza, provides an insightful depiction of such an identity trap in the context of a Muslim neighbourhood of Mumbai during a period of heightened religious tensions.

Ortega, J., & Tangerås, T. (2008). Unilingual versus bilingual education: a political economy analysis. *Journal of the European Economic Association, 6*(5), 1078–1108.
Radhakrishnan, M. (2001). *Dishonoured by history: "Criminal tribes" and British Colonial Policy*. Delhi: Orient Blackswan.

Private Giving in Higher Education

Parimal Bag and Debasis Mondal

1 Introduction

In India, failure of private giving to higher education is the norm. In particular, education at all levels is almost all publicly funded. The supply of good quality educational institutions is thus severely limited when viewed in relation to the huge demands by the country's large population base at the young age group. Provision of education on profit motives is unlikely to be supported by the political class and the general public. Even with recent deregulation entry of non-profit private universities are very few and, to our knowledge, there is no documentation of noticeable initiatives by philanthropists. The end result is a sorry tale of pent-up demands.

The absence of philanthropists in Indian education sector provides a striking contrast when compared with the USA. Many of the top quality universities in the USA are private. While a large part of the costs of private education is borne by student fees, a substantial part of funding of US universities come in the form of big endowment donations by rich philanthropists and alumni.

A recent online report (dated June 28, 2017), "Giving USA 2017: Implications for Higher Education" has the following[1]:

[1]Source: Giving USA 2017: The Annual Report on Philanthropy for the Year 2016, published by Giving USA Foundation, a public service initiative of The Giving Institute, researched and written by the Indiana University Lilly Family School of Philanthropy. See http://info.jgacounsel.com/blog/giving-usa-2017-implications-for-higher-education.

P. Bag (✉)
Faculty of Arts and Social Sciences, Department of Economics,
National University of Singapore, AS2 Level 6, 1 Arts Link, Singapore 117570, Singapore
e-mail: ecsbpk@nus.edu.sg

D. Mondal
Department of Humanities and Social Sciences, Indian Institute of Technology Delhi,
Room no - MS 606 Hauz Khas, New Delhi 110016, India
e-mail: debasis@hss.iitd.ac.in

P. Ray et al. (eds.), *Economics, Management and Sustainability*,
https://doi.org/10.1007/978-981-13-1894-8_7

> *Giving USA* estimates that giving to education organizations (of which giving to higher education accounts for 70% of the total) increased 3.6% in 2016 to $59.77 billion. The Council for Aid to Education's *Voluntary Support of Education* (VSE) survey shows that individual giving (alumni and non-alumni) to education declined by 7.4%, while contributions from corporations and foundations grew by 14.8% and 7.3%, respectively. That said, giving from individuals still comprises the largest source of funds given.
>
> Over the past two years, Giving USA estimates that giving to education grew 12.5%, which was surpassed only by the growth in giving to international affairs (20.7%) and environment/animals (13.5%).
>
> According to the VSE study, alumni giving accounted for 24.2% of total giving to higher education institutions in 2016, a smaller proportion than the previous year. However, this decline comes on the heels of extraordinary growth in individual giving in 2015 (10.2% from alumni and 23.1% for non-alumni).
>
> Giving to education continues to be the most popular philanthropic cause for high net worth donors, according to the Wealth-X and Arton 2016 report, Changing Philanthropy: Trend Shifts in Ultra Wealthy Giving. This report notes that 47% of donations from high net worth individuals reportedly went to education, with 22% supporting higher education specifically. These results continue to bode well for higher education institutions.

University World News (http://www.universityworldnews.com/article.php?story=20110204222722977) runs the headline (dated February 06, 2011), "INDIA: Charity not beginning at home for universities" with excerpts such as the following:

> It is interesting that many of the Ivy League colleges in the US have received donations from their Indian alumni. But the same alumni who have made it big in life because of an Indian education have failed to give back to their home institutions, said Prabhat Lal, professor of international relations at Jawaharlal Nehru University in New Delhi.
>
> Philanthropic donations to Indian institutions are not completely absent but they are few and far between and pale in significance beside the huge donations, often by the same donors, made to western universities.

In this paper, we will view education as a consumption good but to be provided only voluntarily. The greater the collective contribution to education, the better the quality of institutions where young people can gather knowledge that serves them not only for future careers but also in enrichment of life experience. We will abstract from the career angle and focus on the utility enhancing part of consumption of education as a public good. Side by side education, there is also a private good that people buy to maximize their utility. The choice a typical consumer thus faces is how much to spend on the private good and how much to contribute to education as the public good. In an economy with relatively fewer population base, one might expect the supply of the voluntarily provided public good not to be that high. But it is also known that free riding in public good contribution generally increases with the population size (Andreoni 1988). Comparing a large country such as India, in terms of population size, and a "small" country such as the USA, how should voluntary provision of education compare is a theoretically interesting question. We provide an explanation of the different norms of giving in higher education in the two countries—greater generosity in the USA and almost negligible private initiatives in India.

Our main idea can be explained by an extension of Krugman (1979). Krugman had argued in an economy with only private goods, consumers with preference for

product variety (as in Dixit and Stiglitz 1977), economies of scale in production and monopolistic competition, how different regions have a tendency to merge into a single conglomerate region. A merged region will sustain more private goods, thus serving consumers' preference for greater product varieties. We consider an economy with one private good and one public good. Then we flip Krugman's formulation to suggest that a larger region will bring down the unit cost of production of the private good due to supplies of many cheaper inputs.[2] This makes private good in the consumption basket more attractive compared to a public good (i.e. education). In contrast, high labour cost due to low population base makes private good more costly and public good relatively more attractive in the USA. In India, people are thus happier consuming the cheaper private good while the USA compensates by having their children educated in high-quality institutions. This, in a nutshell, is the key message of this paper.

In terms of formal results, we show the following. First, if national border control were not present so that people could move freely between countries, and production technologies in the two countries were uniform, there need not be concentration of population in a single country, i.e. agglomeration might not happen, as suggested by Krugman (1979). Instead, people might live separately in the two countries with uniform per capita welfare as the equilibrating force, as a special case of Proposition 2. As part of this process, the smaller country (USA in our example) may produce more of the public good. If one were to force agglomeration, per capita welfare would come down even with private good becoming cheaper because voluntary provision of the public good will suffer. Endogenous segregation of regions despite increasing returns to scale in the private good's production and necessarily no other exogenous differences is a new and surprising result.

In fact, we show a range of possibilities: increases in population size may increase the equilibrium level of public good, lower it, or sometimes produce a non-monotonic effect by first increasing the level and then lowering it, while the cost of the private good always declines. As a result in some situations per capita welfare monotonically increases with population size, while in others per capita welfare exhibits an inverted-U pattern initially rising and then falling. Welfare may even initially increase, then decrease and eventually may rise again (Proposition 1, Figs. 1 and 2 and Table 1). The broad implication is that with free mobility of labour across countries (that are not integrated through trade, e.g. due to prohibitive trade/transportation costs), one may see either agglomeration or segmentation. Encompassing the alternative possibilities of agglomeration and segregation in one model enriches our understanding of migration similar to Tiebout models.

We also add couple of insights that we hope will improve our understanding of agglomeration. Specifically, an important issue in Krugman's (1979) discussion involved how the initial distribution of population among different regions might influence which region attracts the population. For example, as long as labour

[2] In our formulation large population impacts on the technology side resulting in low-cost supply of the single private good, similar to how large population enables supply of many product varieties in Krugman's economy working via the consumers' love-for-variety preference route.

productivity is the same everywhere, regions with the same population size would have no difference in welfare. However, if there is any difference in productivity, migration may lead to the wrong outcome. Consider, for example, a world with only two regions that are identical everywhere except that in one region private goods are more costly to produce. Then it is clearly desirable that all labour should move to the less costly region. But if the higher-cost region starts with a large population base, per capita welfare will be higher in that region drawing population towards it. Thus, in a model with only a private good, unless there are other exogenous differences, initially uneven distribution of population is the only reason for inefficient agglomeration (Proposition 3). However, in our model with both private good and a public good, even with identical distribution of initial population *the more costly country might offer a better welfare* prospect to cause migration to move in its direction (Proposition 4).[3] This paradoxical result is easier to understand if one considers the improved appeal of the public good as the private good becomes more expensive. The utility loss from higher price of the private good could be more than made up for from higher quality (or quantity) of the public good.

The migration flow towards the inefficient region need not be a pathological outcome from wrong initial population distribution (including identical distribution). In fact, if one were to move the entire population to the less costly (in terms of private good's production) region, per capita welfare could go down. In this sense, inefficient agglomeration could be welfare-dominant rather than welfare-immiserizing.

While we develop the analysis based on Krugman (1979), two papers have been instrumental in thinking about the main ideas explored in this paper. Pecorino's (2009a) is the first analysis of public good in a general equilibrium setting in a variant of Krugman's monopolistic competition model, studying specific conditions under which more population will lead to reduced voluntary provision.[4] Pecorino, however, does not study Krugman's original question of agglomeration.

Equally relevant is Epple and Romano (2003). The authors study voluntary provision of a public good along with dual provision (i.e. voluntary and tax-financed provision). For special type of public goods supported only by a minority of voters, it is shown that only voluntary provision can arise in equilibrium (Corollary to their Proposition 5). For voluntary provision, the examples cited are private funding of US medical research, private and corporate contributions to art museums, and historic preservation societies. Funding of higher education, strictly speaking, is not entirely by voluntary contributions because students are also charged tuition fees. However, to isolate the incentives for voluntary giving we view higher education as a consumable pure public good financed by large endowment gifts by corporate donors, philanthropists and alumni. Certainly major initiatives by rich private universities in the USA in hiring top quality faculties, setting up science and medical laboratories,

[3]We refer, like Krugman (1979), a region to be *inefficient* relative to another region if marginal cost of private good production is higher in the first region. Thus, efficiency is a slightly restricted concept as it does not consider the broader benchmark of social optimality with the addition of the public good.

[4]Pecorino (2009b) analyses the effect of group size on public good in a much simpler economy without production but allowing for rivalry in public good's consumption.

starting a new department, etc., are made possible by big fund drives where universities are known to tap into potential big donors who are likely to support higher education. We take the view that these initiatives improve both the quality of education and its level, benefitting all citizens as stakeholders of higher education and learning.

The rest of the paper is organized as follows. In Sect. 2, we present the basic model. The main analysis of the public good economy is contained in Sect. 2.1, with Conclusion in Sect. 3. An Appendix contains a proof.

2 Preliminaries

There are two types of good in the economy—a pure public good, and a private good. The public good is produced using labour with a one-to-one technology that produces one unit of output using one unit of labour. The private good is produced using n number of different intermediate inputs. These inputs are combined using a CES aggregator to get the final output. We assume that final goods are produced in a perfectly competitive market structure while the market structure for the intermediate inputs is monopolistically competitive.[5]

There are L individuals (or consumers) who each inelastically supplies one unit of labour, earns a competitive wage w and spends it on the private good and contribution towards the public good. Denoting g_j to be the dollar contribution towards the public good by consumer j, $G = \sum_{j=1}^{L} g_j$ is the total voluntary contribution by L consumers. G/w is the total amount of public good produced, with one unit of labour translating into one unit of public good.

The consumers have identical preferences summarized by a very general form of *CES* utility function (see Chap. 1 of Varian 1992) allowing for a broad range of substitutability between the public good and the private good. Representative consumer j solves:

$$\max_{x_j, G} \quad U_j = \left[\eta x_j^r + (1-\eta)G^r\right]^{\frac{1}{r}}, \quad \text{where} \quad 0 \neq r \leq 1, \quad \eta \in (0,1), \tag{1}$$

$$\text{subject to} \qquad p_x x_j + G = w_j + G_{-j}, \quad \text{for all} \quad j = 1, ..., L. \tag{2}$$

Consumption of the private good by person j is given by x_j and its (per unit) price is denoted by p_x. The price of the public good is normalized at unity. We define the contribution by all people except j as $G_{-j} \equiv \sum_{k=1, k\neq j}^{L} g_k$ where g_j is the contribution by person j only. The parameter η in utility function (1) measures the preference bias for the private good.

[5]The basic structure of our model builds on our previous work on group size paradox, Bag and Mondal (2014), and Pecorino (2009a), both of which in turn had built on Krugman (1979).

We define the elasticity of substitution between the private good and the public good as $\epsilon = \frac{1}{1-r} \geq 0$. Setting the Lagrangian for the problem of maximizing (1) subject to (2), we derive the following first-order conditions:

$$\frac{1}{r}\left[\eta x_j^r + (1-\eta)G^r\right]^{\frac{1}{r}-1} \eta\, r\, x_j^{r-1} = \lambda_j\, p_x, \tag{3}$$

$$\frac{1}{r}\left[\eta x_j^r + (1-\eta)G^r\right]^{\frac{1}{r}-1} \eta\, r\, G^{r-1} = \lambda_j. \tag{4}$$

Equations (3) and (4) equalize the marginal benefit and cost of consuming private and public goods, respectively. The marginal utility of income of person j is denoted by the Lagrangian multiplier $\lambda_j > 0$. Using Eqs. (3) and (4), we can write the demand for private good by agent j as

$$x_j = G\left(\frac{\eta}{1-\eta}\frac{1}{p_x}\right)^{\frac{1}{1-r}}.$$

Summing over all individuals, the aggregate demand for the private good is given by

$$X = \sum_{j=1}^{L} x_j = L\, G\left(\frac{\eta}{1-\eta}\frac{1}{p_x}\right)^{\frac{1}{1-r}}. \tag{5}$$

Using Eq. (2) and the definition of G_{-j}, the aggregate budget constraint for the entire economy can be written as $p_x \sum_{j=1}^{L} x_j + G = \sum_{j=1}^{L} w_j$. Since all individuals are alike, they earn the same wage, w. Therefore, using this aggregate budget equation and Eq. (5), we solve for the aggregate public good as,

$$G = \frac{w}{\left(\frac{\eta}{1-\eta}\right)^{\frac{1}{1-r}} p_x^{-\frac{r}{1-r}} + \frac{1}{L}}. \tag{6}$$

As is expected, the aggregate amount of public good is strictly increasing in the income/wealth of the individual (w) and in the size of the contributing group (L). The price of the private good affects G in the standard way depending upon the elasticity of substitution parameter r. For the Cobb–Douglas case with $r = 0$, the public good is invariant to any change in the price of the private good. This completes the description of the demand side of the problem.

Turning to the supply side, we assume that the public good is produced using a one-to-one production technology as given by

$$L_G = G, \tag{7}$$

where L_G is the amount of labour used in producing the public good. Since the public good is produced competitively, and we normalize its price at unity, the wage rate is determined as

$$w = 1. \tag{8}$$

This implies that everybody earns an income equalized to unity and aggregate income is equal to L in the economy.

The private good is produced using n intermediate inputs as per the following CES production technology:

$$X = \left(\sum_{i=1}^{n} y_i^{\theta} \right)^{\frac{1}{\theta}}, \quad \text{where } \theta \in (0, 1). \tag{9}$$

Define $\sigma = \frac{1}{1-\theta}$ to denote the elasticity of substitution among intermediate inputs in the final goods production function. Each intermediate input, y_i, is produced using labour and with the help of a linear production technology as given by

$$l_i = \alpha + \beta y_i, \quad i = 1, ..., n,$$

where α is the fixed labour requirement and β is the marginal labour requirement. Therefore, it takes l_i units of labour to produce y_i units of the ith intermediate input.

The market structure for the final goods are assumed to be perfectly competitive. A final good producer maximizes profit by solving the following problem:

$$\max_{\{y_i\}} \ \pi_X = \ p_x X - \sum_{i}^{n} p_{y_i} y_i \quad \text{subject to Eq. (9)},$$

where p_{y_i} denotes the price per unit of the ith intermediate input. This exercise generates the following demand for the ith intermediate input,

$$y_i = \frac{p_{y_i}^{-\frac{1}{1-\theta}} (p_x X)}{\sum_{i=1}^{n} p_{y_i}^{-\frac{\theta}{1-\theta}}}, \quad i = 1, ..., n. \tag{10}$$

Using Eqs. (9) and (10), we get the following expression of the price of the final private good in terms of price of the intermediate inputs,

$$p_x = \left(\sum_{i=1}^{n} p_{y_i}^{-\frac{\theta}{1-\theta}} \right)^{-\frac{1-\theta}{\theta}}. \tag{11}$$

Each intermediate input producer acts like a monopolist in its own market segment. The profit maximization problem of the ith intermediate input producer can be written as

$$\max_{\{p_{y_i}\}} \ \pi_{y_i} = \ p_{y_i} y_i - w(\alpha + \beta y_i) \quad \text{subject to Eq. (10)}.$$

We assume the presence of large number of intermediate input producers so that each individual producer ignores the effect of others' price on its own demand. Specifically, the term, $\sum_{i=1}^{n} p_{y_i}^{-\frac{\theta}{1-\theta}}$, in Eq. (10) is treated as a fixed value in the above profit maximization. Under this assumption and using Eq. (8), the profit maximizing price of ith intermediate input takes the following form:

$$p_{y_i} = \frac{\beta}{\theta}, \qquad \text{for all } \; i = 1, ..., n. \tag{12}$$

Therefore, intermediate input prices become a constant mark-up over the marginal cost of production and this does not vary with i, given the identical production technology assumption. Henceforth, we get rid of the subscript i.

Free entry into the intermediate input production market implies zero profit. Setting $\pi_y = 0$ and using Eq. (12), the equilibrium supply of ith intermediate input is given as

$$y_i = \frac{\alpha\theta}{(1-\theta)\beta}, \qquad \text{for all } \; i = 1, ..., n. \tag{13}$$

Let us denote the aggregate demand for labour in production of intermediate inputs as $L_y = \sum_{i=1}^{n} l_i$. Then, using Eqs. (12) and (13), we get the following expressions,

$$L_y = \frac{n\alpha}{1-\theta}, \tag{14}$$

$$X = n^{\frac{1}{\theta}} \frac{\alpha\theta}{(1-\theta)\beta}, \tag{15}$$

$$p_x = n^{-\frac{1-\theta}{\theta}} \frac{\beta}{\theta}. \tag{16}$$

Finally, equilibrium in the labour market (i.e. $L_G + L_y = L$) gives us the following solution for the number of intermediate inputs:

$$n = \frac{(1-\theta)(L-G)}{\alpha}. \tag{17}$$

2.1 Solution

Using Eqs. (6), (8), (16) and (17), we get the following equilibrium condition involving G[6]:

$$LG = (L-G)^{\frac{\theta-r}{\theta(1-r)}}\Omega, \tag{18}$$

[6]By Walras' law, it can be shown that aggregate demand for private good is equal to its aggregate supply in the equilibrium (i.e. value of X from Eq. (5) is equal to the value of X from Eq. (15)).

where the expression Ω is a positive constant with $\Omega \equiv \left(\frac{\alpha}{1-\theta}\right)^{\frac{r(1-\theta)}{\theta(1-r)}} \left(\frac{\beta}{\theta}\right)^{\frac{r}{1-r}} \left(\frac{1-\eta}{\eta}\right)^{\frac{1}{1-r}}$. We assume $1 > \theta \geq r$ (i.e. $\sigma \geq \epsilon$) to guarantee the existence of a unique solution of G from Eq. (18). This assumption implies that intermediate inputs in the production function are more substitutable among each other than finals goods are in the utility function. From now onwards, we will interchangeably use the notations σ and ϵ in place of θ and r, respectively.

Define the unique solution of G from Eq. (18) as $G^*(L)$. It can be shown that,

$$\frac{dG^*(L)}{dL} > (=; <)\ 0 \quad \text{iff} \quad G^*(L) > (=; <)\ \left(\frac{\epsilon - 1}{\sigma - 1}\right)L. \tag{19}$$

When $\epsilon \leq 1$, public good is increasing in population size. For $\epsilon > 1$, $G^*(L)$ curve has a global maximum at $L = \hat{L}$. For all $L < (>)\hat{L}$, G^* increases (decreases) with L. We are now interested to see the welfare properties of this model economy as a function of its population size and other cost parameters related to the production of the private good.

3 Welfare Effects

When the equilibrium level of final private good and the level of public good both increase with population size, welfare must increase. The equilibrium number of input varieties (denoted by n) always increase with population size irrespective of the value of ϵ.[7] When $\epsilon \leq 1$, G^* is strictly increasing in L, and hence, welfare must go up with population size in this case. For $\epsilon > 1$, G^* will decrease for all $L > \hat{L}$, so there is a possibility that welfare might decrease in L for some ranges of L. To show this formally, using Eqs. (5), (16), (17), (18), we rewrite the utility (or welfare) of a representative agent in (1) as follows:

$$U_j = \left(\frac{L - G^*}{LG^*} + 1\right)^{\frac{1}{r}} G^* \ (1 - \eta)^{\frac{1}{r}}. \tag{20}$$

The expression, $\frac{L-G^*}{LG^*}$, in the right-hand side of the above expression always rises with L. To see this, just rewrite Eq. (18) as $\frac{L-G}{LG} = \frac{1}{\Omega}\ (L - G)^{\frac{r(1-\theta)}{\theta(1-r)}}$; and note that $(L - G)$ is an increasing function of L (see footnote 7). Thus as L increases, welfare goes up due to the (certain) reduction in price of the final private good, p_x. The possibility that welfare goes down with an increase in L arises only when G^* decreases with L. We

[7] To see this, note that the expression $\frac{dG^*(L)}{dL}\frac{L}{G^*}$, which is that elasticity of G^* w.r.t. L, will always lie between -1 and $+1$. Hence, the expression LG^* will always increase due to an increase in L even if G^* falls due to a rise in L. But this implies that, $(L - G^*)$ will also rise in L from Eq. (18) since $\theta \geq r$ by assumption. This, in turn, implies that n is always an increasing function of L from Eq. (17).

Table 1 Per capita welfare with $\beta = 5$ and $\beta = 10$ ($\sigma = 5,\ \epsilon = 3,\ \eta = 2/5,\ \alpha = 1$)

	Region 1: $\beta = 5$			Region 2: $\beta = 10$		
L	U	p_x	G	U	p_x	G
30	11.304	5.977	24.023	13.695	10.799	29.439
31	11.493	5.833	24.411	14.126	10.461	30.363
32	11.670	5.699	24.770	14.553	10.142	31.279
50	13.225	4.304	27.787	21.510	6.680	46.168
58	**13.326**	3.989	27.890	24.043	5.868	51.567
60	13.321	3.925	27.856	24.616	5.703	52.786
62	13.309	3.864	27.804	25.165	5.549	53.953
63	13.299	3.836	27.773	25.431	5.476	54.516
124	11.628	2.954	23.798	32.804	3.446	69.927
125	11.599	2.946	23.732	32.829	3.431	69.976
126	11.571	2.938	23.668	32.853	3.417	70.02
127	11.542	2.931	23.603	32.875	3.402	70.061
128	11.514	2.923	23.539	32.895	3.388	70.099
144	11.078	2.815	22.56	**33.030**	3.189	70.299
149	10.95	2.785	22.272	33.012	3.137	70.236
150	10.925	2.780	22.216	33.005	3.127	70.217
250	9.003	2.394	17.962	30.394	2.531	64.279
500	6.738	1.989	13.011	23.880	2.029	50.027
1000	5.021	1.666	9.279	17.686	1.677	36.600
10^5	1.293	0.525	0.932	2.475	0.525	3.729
$2 * 10^5$	1.223	0.442	0.659	1.985	0.442	2.635
$4 * 10^5$	1.208	0.372	0.466	1.650	0.371	1.864
$5 * 10^5$	1.214	0.351	0.417	1.569	0.351	1.668
$6 * 10^5$	1.222	0.336	0.381	1.509	0.336	1.522

have seen that for $\epsilon > 1$, G^* is non-monotonic in L. In particular, G^* is decreasing for all $L > \hat{L}$. Hence, welfare may fall with population size at some point in the range exceeding $\hat{L}$.

Given the complicated expression of (20) (as we need to solve for G^*), a neat characterization of per capita welfare with respect to population size is difficult. So for some of the results to follow, we will rely on numerical simulations. By fixing specific values for $(\alpha, \beta, \eta, \sigma, \epsilon)$, the values of U, equilibrium p_x and equilibrium G for various values of L are reported in Table 1. We take two independent regions—region 1 and region 2. They share the same parameter values except for β. Region 1 has a lower β value than region 2. Thus, the price of each intermediate input in region 2 is more than in region 1. This implies that, private good is more costly to produce in region 2 given the same number of intermediate inputs. Nevertheless, as

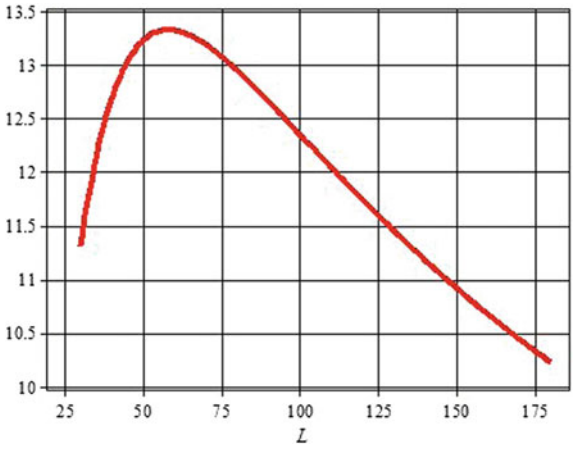


Fig. 1 Inverted-U per capita welfare (Table 1: region 1)

we will see below, region 2 may provide better per capita welfare than region 1. This is because of the larger level of consumption of the public good in region 2.

We report the possibility of an ***inverted-U welfare***. That is, welfare may initially go up and then down as population size increases.

As Table 1 shows, as population size increases, per capita welfare increases initially (in both regions), reaches maximum at $L = 58$ in region 1 (at $L = 144$ in region 2) and then starts falling: an inverted-U per capita welfare shown for region 1 in Fig. 1. However, welfare curve may even take a U-turn for large values of L as shown in Fig. 2 where L varies in the range $[2 * 10^5, 6 * 10^5]$ (and $\beta = 5$).[8] The reason is, for L becoming very large, G approaches zero while n approaches infinity (see Eq. (20) and note that $\frac{L-G^*}{LG^*}$ is proportional to n). Thus, the property that G is bounded below by zero makes it possible that per capita welfare might grow unbounded in L. We can summarize the above discussion as follows:

Proposition 1 *(*Per capita welfare)

(a) Per capita welfare is increasing in L if $\epsilon \leq 1$.
(b) If $\epsilon > 1$, per capita welfare is increasing in L for $L \leq \hat{L}$ but it may fall at some point in the range of $L > \hat{L}$, thus giving rise to an inverted-U pattern.
(c) If $\epsilon > 1$, per capita welfare may exhibit U-turn after the inverted-U pattern as can be seen combining Figs. 1 and 2.

We also show the relationship between region's population size and price of the private good and the level of public good in Figs. 3 and 4. Note that, the costly region (region 2) has higher prices (Fig. 4) but also provides higher amounts of the public good (Fig. 3). This result provides an explanation of why in the USA (respectively

[8] See last two entries in Table 1 (region 1). For $L > 4 * 10^5$, welfare starts rising with L further up. This is represented in Fig. 2. Also note that in Table 1 some of the n-values for region 2 are omitted as those were coming out to be less than 1. The numbers have been generated using the software Maple-12.

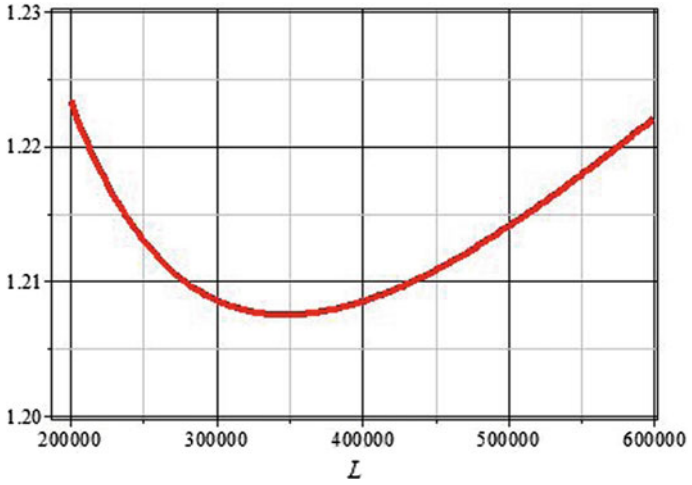
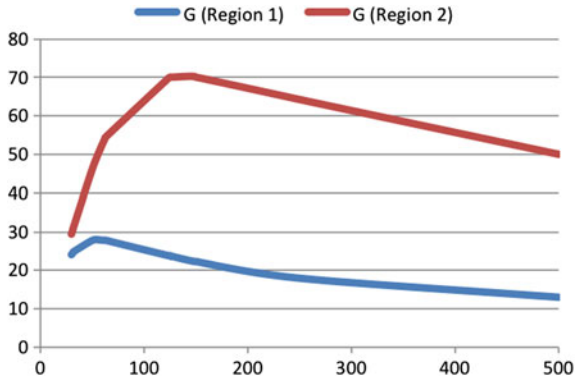


Fig. 2 U-shape per capita welfare for L in the range $\in [2 * 10^5, 6 * 10^5]$ (region 1)

Fig. 3 Costly region with higher provision



India) prices are higher (lower), and at the same time, voluntary giving to higher education is higher (lower). (Also, private good prices in both regions converge to zero as L become very large.)

3.1 Agglomeration or Segregation

Non-monotonic welfare (Proposition 1**(b)** and **(c)**) gives rise to either possibility—agglomeration or segregation of region—depending on total population size. To

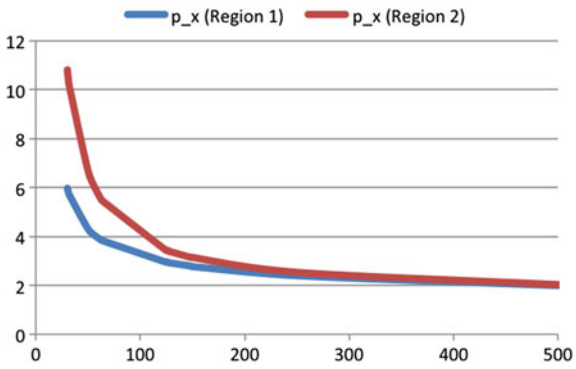


Fig. 4 (Relative) price of the private good higher in region 2

analyse this issue, we focus only on the inverted-U portion of the welfare curve.[9] Let us assume that there are two identical regions each with $L = 30$, the benefits from intermediate input varieties these regions produce are exclusive to the regions, there is no trade, and workers are immobile. From column 2 of Table 1 (region-1) individuals in these regions will enjoy a per capita welfare of **11.304**. Now if we allow free mobility of workers across regions, do we see agglomeration? In this case, the answer is "yes" as the combined region will have a population size of 60, each enjoying welfare **13.321**.

We next ask whether one big region can segregate into several smaller regions. Our answer, once again, is "yes". An example is given following the next proposition. But the intuition is simple: a combined region may be on a downward-sloping part of the inverted-U welfare curve, which means there is always some split of the population size L into two regions, say, such that the welfare is higher in both those regions separately than the combined region. This creates an incentive for the region to break up. However, one has to be careful about how the segregation is going to play out. Segregation has to be welfare improving and migration-proof, the latter being a stability requirement. The following proposition provides a further guidance to the nature of segregation[10]:

Proposition 2 *(*Uniform segmentation) Starting from a single region of size L and assuming an inverted-U welfare curve (i.e. Proposition 1 part (b) applies), suppose there is a break up of the region into n smaller regions with population size in the i^{th} region L_i, $i = 1, ..., M$ and $\sum_{i=1}^{M} L_i = L$. Then,

(a) It must be the case that $L_i = L_j$ for all $i \neq j$ $(= 1, ..., M)$; i.e. breakup must generate identical regions.
(b) The equilibrium number of breakups, M, must satisfy: $U(\frac{L}{M}) \geq \max\{U(\frac{L}{2M}), U(\frac{2L}{M})\}$.

[9]When per capita welfare curve is increasing on the "entire" domain L (as appropriately defined), agglomeration will be the only outcome as predicted in Krugman (1979). However, agglomeration can happen even without such strong requirement.

[10]These are, of course, in the form of necessary conditions.

To illustrate part (b) of Proposition 2, let us take a specific example from Table 1. Suppose initial population size of a region is $L = 1000$. Breaking up into two equal-sized regions is perfectly possible in this case as $U(500) = 6.738 > U(1000) = 5.021$. Each region of size 500 will have a further incentive to break up since $U(250) = 9.003 > U(500)$. We can now apply the result of part (b) of Proposition 2 to determine equilibrium n: it must be of the form $M = 2^k$ for any positive integer k. For $k = 3$, we have $M = 8$ and can check from Table 1 that $U(1000/4 = 250) < U(1000/8 = 125) < \min\{U(\lfloor 1000/16 \rfloor = 62), U(\lfloor 1000/16 \rfloor + 1 = 63)\}$, so $M = 8$ cannot be equilibrium (symbol $\lfloor z \rfloor$ denotes the largest integer contained in z). Next, for $M = 16$ ($k = 4$), we can check that inequality in part **(b)** is satisfied, with eight regions of size 62 and the remaining eight of size 63; this breakup is migration-proof not only against individual deviations but against coalitional deviations as in Conley and Konishi (2002).

Remark 1 In the proof of Proposition 2 in the Appendix, we check migration-proofness by simply comparing per capita welfare in different regions. If welfare differs between any two regions for any given population division, we say that there should be an incentive for some people to move from the low-welfare region to the high-welfare region. For our stability notion (or migration-proofness), the incentive refers to an individual member as opposed to any coalition of members (as in Conley and Konishi 2002). Since the argument we use for equality of regions is in the form of a "necessary condition", looking at individual incentives rather than for arbitrary coalitions serve our purpose.

Remark 2 Our definition of migration-proof equilibrium has the notion stability built in it. The issue of stability is sometimes viewed separately from the equilibrium notion in partial equilibrium models with local public goods. For example, see Hindriks and Myles (2013, ch. 7).

3.2 Inefficient Agglomeration

With intermediate input varieties and only a private good, i.e. in a ***Krugman's world***, agglomeration is a natural outcome: one big region allows more firms to enter (due to scale economies), producing a greater variety of intermediate inputs and catering to a greater number of consumers, thus improving per capita welfare.[11] But which region ends up with the population depends on the initial distribution of population. As long as labour productivity is the same everywhere, regions with the same population size

[11] According to Krugman's own word (see Krugman 1979, p. 478): "*In the presence of increasing returns factor mobility appears to produce a process of agglomeration. If we had considered a many-region model the population would still have tended to accumulate in only one region, which we may as well label a city; for this analysis seems to make most sense as an account of the growth of metropolitan areas*".

would have no difference in welfare among them. However, if there are regional differences in productivity, which region attracts the population will make a difference to people's welfare—the process of migration may lead to the wrong outcome. Consider, for example, a world with only two regions that are identical everywhere except for the fact that the variable costs of production (of intermediate inputs) are higher in one region while fixed costs are the same. Then it is clearly desirable that all labour should move to the other (efficient) region. But if the higher-cost/inferior region starts with a large enough share of the population, per capita welfare could be higher in that region prompting migration to proceed towards the inefficient region. But if the two regions have the same initial population, the inefficient region will have a worse initial per capita welfare, prompting migration to move towards the efficient region. These observations that follow directly from Krugman (1979) are summarized below:

Proposition 3 *(*Agglomeration in Krugman world) In Krugman's world, agglomeration is the logical outcome if one allows for free mobility of labour. Further, the following observations can be made:

(a) Starting from two identical regions in terms of population size, tastes and technologies so that initial per capita welfare is the same, agglomeration is always welfare-neutral, i.e. to which region the population moves will not matter.
(b) Starting from two regions with identical population and tastes but one being the higher-cost region, migration will always move towards the more efficient region: agglomeration is welfare-dominant.
(c) Starting from two regions with identical tastes, if a higher-cost region (due to higher β or α or both) has much higher population initially, then the migration may move towards this inferior region: agglomeration is inefficient and welfare-immiserizing.

One notable contrast of our model is that, even if we start with an *identical* distribution of initial population, inefficient region might yield higher welfare initially and may offer a better end-welfare prospect following agglomeration. This is purely due to the presence of the voluntarily provided public good.[12] Lower productivity in private intermediate goods production (i.e. higher values of β and/or α) either leads to higher prices or fewer varieties (or both). Both lead to an increase in the price of the final private good.

For *gross substitutability in preferences* ($\epsilon > 1$), demand for the public good goes up and the final private good goes down. Collective higher provision of the public good and lower private good consumption (with fewer input varieties) will impact on per capita welfare in opposite directions. As a result, it may well happen that the inefficient region ends up providing, initially, a higher per capita welfare, triggering migration and attracting all the population. What is even more surprising is that even if one ignores the initial trigger for the migration and simply moves the entire

[12]In a model without public good, any increase in prices due to inferior technology and possibly fewer intermediate input varieties should result in a lower welfare.

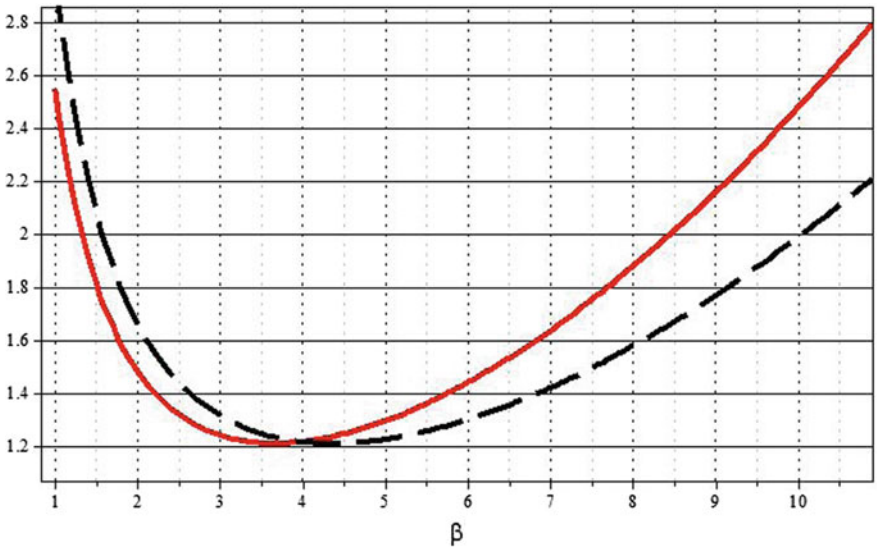


Fig. 5 U-shape welfare against marginal cost parameter β [$L = 10^5$(red curve), $L = 2 * 10^5$(black dashed curve), $\sigma = 5$, $\epsilon = 3$, $\eta = 2/5$, $\alpha = 1$]

population from inefficient to efficient region, per capita welfare might actually drop. The usual economic logic that if a region's productivity were to improve, then people in that region should be better off (given the representative agent assumption), may fail to hold. Below we illustrate this possibility using numerical simulations.

Consider once again Table 1. Marginal cost of intermediate inputs for region 2 is twice the marginal cost for region 1: $\beta = 10$ versus $\beta = 5$. Per capita welfare (U) in region 2 is higher for all reported population sizes, giving rise to the possibility of migration tipping towards the higher-cost, region 2. (We also report corresponding values of p_x and G to highlight the trade-off between p_x and G.) Now to illustrate the possibility that agglomeration in region 2 can be welfare-dominant (relative to region 1), let us go beyond Table 1 and consider Fig. 5. Here welfare curves are plotted for a continuum of β-values.[13] Looking at the rising part of the welfare curve in colour red, consider regions 1 and 2 both of size $L = 10^5$ and the β's as specified earlier ($\beta = 5$ for region 1 and $\beta = 10$ for region 2). Initially, in region 1, $U \approx 1.29$ while for region 2, $U \approx 2.47$. This will move people from region 1 to region 2 resulting in per capita welfare $U \approx 1.98$ for the total population $L = 2 * 10^5$ (refer the welfare curve in colour black (dashed curve)). However, if the same population were to be placed in the low-cost region 1, per capita welfare would have dropped to barely above $U = 1.2$ (refer the black-coloured dashed curve against $\beta = 5$).

[13] Though in Table 1 welfare is higher for the higher β, clearly this is not always the case; for $\epsilon > 1$, there is a trade-off between higher G and lower product variety arising from lower productivity. In Fig. 5, we show U-shaped welfare against β.

Next, for *complementarity in preferences* ($\epsilon < 1$), welfare always goes down due to an increase in price of the final good. This is because, higher final good's price would generate lower demand for both the final private good and the public good. Here one gets back Krugman's result; that is, with identical population distribution, inferior region would generate lower per capita welfare. For the *Cobb–Douglas case* ($\epsilon = 1$), G is unaffected to any change in the price of the final private good and hence per capita welfare would be lower in the high-cost/inefficient region.

We summarize the above observations in the following proposition:

Proposition 4 *(*Welfare-dominant inefficient agglomeration) Consider a world with CES utility and increasing returns in the production of the private goods as specified in Eqs. (1) and (9), intermediate inputs markets monopolistically competitive and public good voluntarily provided.

Starting from two regions with identical population sizes but differing technologies, any agglomeration will exhibit one of following characteristics:

(a) Given complementarity in preferences, i.e. $\epsilon \leq 1$, the technologically inferior region will always offer a lower per capita welfare initially. In such a case, agglomeration will be tipped towards the efficient region and will also be welfare-dominant.
(b) Given substitutability in preferences, i.e. $\epsilon > 1$, the technologically inferior region may generate, initially, a higher or lower per capita welfare. In the case of the former, agglomeration will be tipped towards the inefficient region but it can be welfare-dominant.

Proposition 4 is an interesting demonstration of the economics of second-best. Two regions are compared which are identical in size, but one having worse technology of private good production. Inefficient production raises the price of the private good, and under substitutability raises provision of the public good. Given that the market equilibrium cannot provide the public good efficiently, it is indeed possible that welfare can be higher in the inefficient region.[14]

As already noted earlier, part **(b)** result of Proposition 4 does not conform to standard logic. The intuition is that more costly region, via its market process (fewer input varieties and high price of the final private good), makes higher voluntary provision of the public good sustainable that would not be possible in the low-cost region, and this public good effect sometimes comes to dominate. However, the result is *not* true generally. To see this, refer Fig. 5 and consider two β values: $\beta = 2.5$ (efficient) and $\beta = 6$ (inefficient). Welfare along the red curve is higher at $\beta = 6$ than at $\beta = 2.5$, i.e. inefficient region offers higher per capita welfare with a population of $L = 10^5$ compared to a similar sized region but with $\beta = 2.5$. Let us check the per capita welfare if all population moves to the inefficient region rather

[14]That lower efficiency in production (i.e. higher opportunity cost) can ultimately lead to higher welfare has also been demonstrated in pure voluntary contribution setting for international public goods. For example, Ihori (1996) has shown in a two-country model that when the price of giving differs, the country with the higher price of donation may enjoy higher welfare.

than the efficient region. Looking at the black dashed curve, it is clear that the entire population should move to the efficient region as welfare at $\beta = 2.5$, ($U \approx 1.45$), is higher than the welfare at $\beta = 6$ ($U \approx 1.3$). Here, initially higher per capita welfare in the inefficient region triggers a movement that is welfare-immiserizing just like in Krugman (1979). If one were to apply the demanding, coalition-based notion of migration-proofness as in Conley and Konishi (2002), then in this example the *entire population* should relocate to the efficient region after the initial process of migration.

■ **Agent heterogeneity**. In our representative agent model, the Nash equilibrium contribution of the public good is solved using symmetry of agent incomes (more precisely, labour endowment) and preferences. It is possible to allow income heterogeneity by considering two income classes, say the rich and the poor, where the rich contributes to the public good and the poor free rides similar to standard public good models without production. The incentives for agglomeration or segregation will have to be addressed more carefully; however, as the usual tension between the rich and the poor will resurface. While the rich and the poor will clash over the public good's provision, the two groups will pull in the same direction with regard to the private goods: more population via Krugman-type agglomeration will tend to raise agents' utilities due to greater intermediate input varieties. Our analysis here should be useful for a more elaborate modelling of this interesting issue.

4 Conclusion

Our modelling of voluntary contribution to higher education as a public good in a general equilibrium setting with identical agents differs from the literature's partial equilibrium analysis. This way we are able to bring in a production sector explicitly where endogenously determined price of the private good impacts on the price of voluntary giving.

We do not model strategic aspects of donation in its full richness because of many agents and the general equilibrium structure. Also, we have remained silent on the role of fundraising drives in generating donations. In the USA, private universities actively solicit donations from alumni, corporate organizations and philanthropists. As Andreoni and Payne (2013) have discussed, people generally do not give unless asked explicitly by charities. So one reason private universities in the USA are so successful in receiving donations are due to their big fundraising initiatives. In this way the USA is unique. To our knowledge, Indian universities make no such major initiatives. This difference in approach in the two countries can of course explain part of the differential giving, besides the theoretical explanation we have offered here.

Last but not least, it might be argued that the Americans donate more simply because they are more wealthy. This argument is not entirely accurate. Big donations in higher education come mainly from rich alumni and philanthropists. The capacity of giving should not be any different whether one considers rich Americans or rich

Indians. Generosity is not to be measured in terms of absolute dollar donations. Rather how much one gives relative to one's wealth should be the correct indicator.

Acknowledgements The first author is grateful to Anup Sinha for his advice and guidance as a teacher. This work is a token appreciation of his generosity. The research reported here received funding from the Singapore Ministry of Education.

Appendix

Proof of Proposition 2. **(a)** We assume L to be a continuous variable.[15] In equilibrium, all regions must enjoy the same level of welfare and there must not be any incentive for people to move from one region to another. Contrary to our claim suppose, there are two regions, i and j, such that $U(L_i) = U(L_j)$ and yet $L_i \neq L_j$ with $L_i > L_j$. Let welfare be maximized on the inverted-U curve at $L = \tilde{L}$. It must then be that $L_i > \tilde{L} > L_j$; all other possibilities can be ruled out given that $U(L_i) = U(L_j)$, $L_i \neq L_j$ and $U(L)$ is maximized at $\tilde{L}$ on the inverted-U curve. But then $U(L)$ is increasing between L_j and $\tilde{L}$. This implies that there is an incentive for some people to move from region i to region j, which is a contradiction.

(b) Suppose after the breakup, there are n identical regions. Using the result in part **(a)**, each region will have $\frac{L}{n}$ population. That any further breakup is not possible requires $U(\frac{L}{2M}) < U(\frac{L}{M})$. Also, no two regions should have the incentives to merge together: $U(\frac{2L}{M}) < U(\frac{L}{M})$. Combining these two inequalities prove this part. **Q.E.D.**

References

Andreoni, J. (1988). Privately provided public goods in a large economy: The limits of altruism. *Journal of Public Economics*, *35*, 57–73.

Andreoni, J., & Payne, A. A. (2013). Charitable giving. *Handbook of Public Economics, Chapter*, *1*, 1–50.

Bag, P. K., & Mondal, D. (2014). Group size paradox and public goods. *Economics Letters*, *125*, 215–218.

Conley, J., & Konishi, H. (2002). Migration-proof Tiebout equilibrium: Existence and asymptotic efficiency. *Journal of Public Economics*, *86*, 243–262.

Dixit, A. K., & Stiglitz, J. E. (1977). Monopolistic competition and optimum product diversity. *American Economic Review*, *67*, 297–308.

Epple, D., & Romano, R. (2003). Collective choice and voluntary provision of public goods. *International Economic Review*, *44*, 545–572.

Hindriks, J., & Myles, G.D. (2013). *Intermediate Public Economics* (2nd ed.). The MIT Press.

Ihori, T. (1996). International public goods and contribution productivity differentials. *Journal of Public Economics*, *61*, 139–154.

[15] For L treated as integers, segregation might not precisely equalize populations across regions. The maximum difference between any two regions' equilibrium population size should be 1.

Krugman, P. R. (1979). Increasing returns, monopolistic competition and international trade. *Journal of International Economics*, *9*, 469–479.
Pecorino, P. (2009a). Monopolistic competition, growth and public good provision. *Economic Journal*, *119*, 298–307.
Pecorino, P. (2009b). Public goods, group size, and the degree of rivalry. *Public Choice*, *138*, 161–169.
Tiebout, C. (1956). A pure theory of local expenditures. *Journal of Political Economy*, *64*, 416–424.
Varian, H. R. (1992). *Microeconomic analysis* (3rd ed.). W.W: Norton & Company Inc.

Being Out of Work: An Analysis of Unemployment and Its Duration in India

Mousumi Dutta and Zakir Husain

1 Introduction

Unemployment refers to a state of not being engaged in any economic activity and seeking work. Though the value placed on employment depends on the type of employment, it is widely acknowledged that unemployment is 'bad'. There are several reasons for this.

Nordhaus (1975) observes that conditions of slump and high unemployment may reduce working hours and overtime opportunities even if the worker is not thrown out of work. Those who are unemployed find it difficult to get a new job, as even opportunities for part-time work become restricted. In such cases, Stewart and Streeten (1971) point out employment creation may be the only mechanism for redistributing income to those who would otherwise remain unemployed.

Unemployment may leave a scar even after the worker regains a new job (Ruhm 1991). Studies in the USA and Britain show post-unemployment earnings losses to be permanent (Gregory and Jukes 2001; Brand 2015; Gangl 2006). Factors responsible for scarring include '… stigma effects of unemployment, loss of workers' firm-specific human capital, human capital depreciation through intensified economic restructuring, and constraints on worker search behaviour' (Gangl 2006). The financial strain associated with being out of work may also affect health status (Marmont et al. 2013). While the reduction in absolute income may reduce family budget on health-related expenditure, the decline in relative income can affect self-esteem and social status, affecting health through psychosocial channels (Tøge 2016).

M. Dutta (✉)
Economics Department, Presidency University, Kolkata, India
e-mail: mousumi.econ@presiuniv.ac.in

Z. Husain
Humanities & Social Sciences Department, Indian Institute of Technology, Kharagpur, India
e-mail: dzhusain@gmail.com

P. Ray et al. (eds.), *Economics, Management and Sustainability*,
https://doi.org/10.1007/978-981-13-1894-8_8

Unemployment and its accompanying economic hardships may create political instability. Nordhaus's theory of political business cycles is based on the assumption that high unemployment levels will lead to a swing in the electorate against the incumbents; this induces parties in power to control unemployment by stimulating the economy just before elections (Nordhaus 1975).

Puritanism views work as valuable, irrespective of its contribution to production. Based on this ethic, writers have underlined the intrinsic worth of employment is intrinsically good, whatever its impact on morale, self-respect and other subjective feelings. This has led researchers emphasising on the demoralising effect of unemployment—'To feel unwanted, not to be able to make any contribution, lowers a man's morale and makes him lose his self-respect' (Stewart and Streeten 1971: 152).

For these reasons, unemployment is treated as one of the important macroeconomic variables, and considerable attention is paid to its measurement. In India, there are two main sources of data on unemployment. They are the decadal Census and quinquennial surveys by the National Sample Survey Office (NSSO). Both these sources have been criticised for their narrow definition of unemployment (Sen 1975). In this paper, we will use a new addition to the NSSO questionnaire (in the 68th round) to show how unemployment can be measured. We will estimate unemployment rates, study its variation over important correlates, identify determinants of unemployment and conclude by a study of the duration of unemployment.

2 Defining Unemployment

2.1 Official Definition of Unemployment

In India data on employment are reported in the decadal Census, Economic Census (for the non-agricultural sector), surveys of the Labour Bureau and NSSO. However, it is only in the NSSO reports and surveys of Labour Bureau that we get estimates of unemployment. The 1961 Census also provides estimates of unemployment.

In NSSO surveys, the definition of unemployment is as follows[1]:

> The activity status is determined by the activity situation in which a person is found during a reference period or at a point of time under reference, which occurs with the person's

[1]The following discussion is largely based on GoI (2001).

> participation in economic and non-economic activities.[2] According to this, a person will be in one or a combination of the following three statuses during a reference period:
>
> (i) working or being engaged in economic activity (work),
> (ii) being not engaged in economic activity (work) and either making tangible efforts to seek 'work' or being available for 'work', if the 'work' is available, and
> (iii) being not engaged in any economic activity (work) and also not available for 'work'.
>
> Activity status (i) above is associated with 'employment', (ii) with 'unemployment' and the last with 'not being in the labour force'. (GoI 2001: 43).

In NSSO reports the concept of usual status is important. This relates to the activity status of a person during the reference period of the year preceding the date of survey. The activity status on which a person spent relatively longer time (major time criterion) during the reference period is considered his or her principal usual activity status. Initially, NSSO used a threefold categorisation of activity status—'employed', 'unemployed' and 'out of labour force'—based on how the respondent had spent the major part of the reference period. From the 50th round, a two-stage dichotomous procedure was adopted. Respondents were initially classified as being 'in the labour force' or 'out of the labour force'; those respondents who were 'in the labour force' were further classified as either unemployed or unemployed.

This is easily understood if we consider the NSSO codes on principal status.

NSSO also has a reference period of the week preceding the date of survey. If the respondent has not been engaged in any economic activity on any of the seven days but was looking for work, (s)he is classified as unemployed. This is called the current activity status and was used in the 11–18th rounds. Subsequently, till the 22nd round, the economic status during the majority of the week was considered in determining whether the respondent was unemployed. From the 27th round, the term current weekly status was introduced.

An approach similar to that of NSSO is followed by the Bureau of Labour. Economic activity is defined as any activity for market-oriented production, agricultural production for self-production and own production of fixed assets. Using a reference period of one year, respondents are classified as workers, unemployed or being out of labour force depending upon how they have spent the majority of the reference period. A second estimate, using a reference period of the week preceding date of survey, is also used to determine the weekly status. Thus, unemployed means:

> Persons, who owing to lack of work, had not worked but either sought work through employment exchanges, intermediaries, friends or relatives or by making applications to prospective

[2]Initially, while referring to 'work', NSSO used the term 'gainful activities'. This referred to any activity adding value to the 'national product' of the country. However, in addition to production of goods and services for exchange, any agricultural production for own consumption and did not go for sale was also considered 'gainful' or 'work'. From the 50th round, NSSO substituted 'economic activities' instead of 'gainful activities'. Economic activity refers to any activity resulting in production of goods and services that adds value to national product is considered as economic activity. It includes production of all goods and services for market, production of primary commodities for own consumption and own account production of fixed assets. This is a slightly broader definition than gainful activities, including own account production of fixed assets.

> employers or expressed their willingness or availability for work under the prevailing condition of work and remuneration are considered as those who are 'seeking or available for work' (or unemployed) (GoI 2012: 16).

Information on economic activity is being collected in the Census since 1872. The 1872 Census, as well as the 1881 Census, also collected information on occupation. In 1891 information on means of subsistence too was elicited. Information on principal and subsidiary occupations of actual workers was collected in the period 1901–1921, with workers being defined as any person earning income.[3] The 1941 Census obtained information on means of livelihood, while the 1951 Census classified the population into dependents and employed (using economic status) and principal and subsidiary means of livelihood. In the 1961 Census, it was decided to classify the population into workers and non-workers. As the category of non-workers comes closest to what we mean by unemployment, this classification merits our attention.

In the case of seasonal activities, any person working regularly for more than one hour daily throughout the greater part of the working season was regarded as a worker. In the case of regular employment in any trade, profession, service, business or commerce a person had to be employed during any of the fifteen days preceding the day on which he was enumerated. If the person was normally working but was absent from work during the reference period due to illness or other cause, was also considered to be a worker. Among others considered to be workers were: trainee or apprentice working with or without wages, public or social service worker engaged in public service activities and political worker engaged in political activities. A person who was offered work but had not actually joined was treated as a non-worker. Adult women engaged in extended System of National Accounts activities, beggars, pensioners, etc., who received income without doing any work were also regarded as non-workers. Thus, the non-worker category includes both persons normally considered to unemployed and those outside the labour force.

From the 1971 Census, the population was classified by main activity. The main activity of a person was ascertained on the basis of how (s)he spent (her)his. For regular work in Industry, Trade or Services the reference period was the week prior to the enumeration; it was one year for seasonal work. In addition, if any person, whether worker or non-worker, made marginal contribution to work this was recorded under secondary activities. From this Census onwards, information on unemployed was not recorded; unemployed were included—along with housewives, rentiers, pensioners, beggars and students—under the category of non-workers.

[3]Thus, even rentiers and pensioners were considered to be workers.

2.2 Sen's Criticism

Sen (1975) starts by pointing out that the basic principle underlying the definition of unemployment in the NSSO reports is the intersection of, what he calls, the income and recognition criteria.[4] This is explained as follows:

> The test, therefore, is based on the *intersection* of two criteria. If one recognises oneself as unemployed and 'seeks work', but regularly does one or two hours of work in the family farm, one does not qualify as 'not working' and therefore has no chance of being taken as unemployed. Similarly if one is 'not working' but not 'seeking work', then again one is not unemployed. The… definition of unemployment covers precisely those who pass *both* the tests (Sen 1975: 120).

Given the stringency of the test, it is not surprising, therefore, that the estimates of unemployment are 'low enough to put may advanced countries to shame' (Sen 1975: 119).

Further, Sen argues, the definitions of both income and recognition criteria are narrow.

For instance, if a person feels that although (s)he is not making any worthwhile contribution, there are no possible sources of work (s)he will not seek work. In that case, under the NSSO definition, the person will be not be treated as unemployed—even though the person falls in the category of unemployed using the recognition approach. 'The interpretation of the recognition approach is, therefore, rather narrow' (Sen 1975: 121). Over time, the addition of the clause 'actively seeking work' to the clause 'making tangible efforts to seek work' had further tightened the recognition criterion.[5]

Similarly, the income approach is also narrowly defined:

> The income approach should, strictly speaking, count as employed only those who would not receive their share of the family income if they stopped working. The test is not whether one is working and receiving an income, but whether one is receiving an income *because* one is working (Sen 1975: 121).

Although this may not make a major quantitative difference, as the distinction between the two is not easily verified empirically, the income criterion does make it difficult for a person to be classified as unemployed, particularly as we are considering only the intersection with the recognition criterion.

Thirdly, NSSO imposes an age restriction. Only persons above 15 years are considered to be in the workforce and can be classified as being either employed or unemployed. Persons below 15 years are not considered unemployed even if they are actively seeking work.[6]

---

[4] 'The income criteria require the person to be earning an income in order to be classified as employed, while the recognition criteria state that the person must have a sense of being engaged in some activity worth his while' (Sen 1975: 5).

[5] Recognising the importance of this point, NSSO has added the clause 'if work is available', thereby widening the recognition criterion.

[6] This criterion is imposed to eliminate child workers from being counted as part of the work force. There is, thus, an ethical dilemma whether we should accept Sen's criticism and treat children

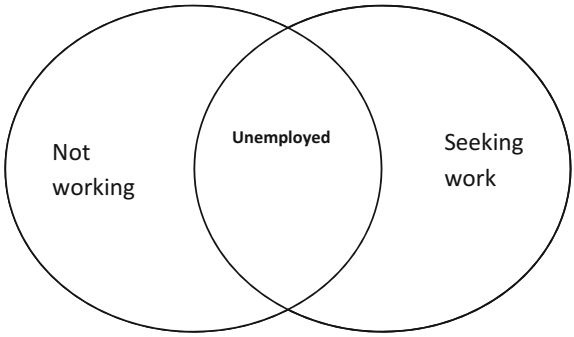


Fig. 1 View of unemployed (*Note* Sen 1975: 120)

2.3 An Alternative Concept

The NSSO elicits information on how many months a worker (i.e. respondent with principal or subsidiary status codes 11–51) has been unemployed. Respondents are also asked whether (s)he had sought work, was available for work and their respective reasons. A comparison of those without work for at least six months with unemployment figures is interesting.

Figure 2 reports the percentage of unemployed workers using the NSSO definition (principal status code of 81). In both rural and urban areas, as well as for both genders, it is very low—well below unity in all cases. However, 1.03% of workers as classified by NSSO have been without work for six months or more (code 11–51); the figure is 2.47% for all persons deemed to be in the labour force (code 11–81). Figure 1 shows the breakup across gender and place of residence for percentage of unemployed workers (as classified by NSSO), percentage of employed workers without work for more than 5 months (labelled 'Without work (Pr. Stat. Worker)') and percentage of persons in labour force without work for at least 6 months (labelled 'Without work').

It can be seen that the discrepancy between the proportion of respondents without work for the greater part of the year preceding the survey and NSSO figures for unemployment is quite large, particularly for females. Further, 3.17% of employed workers were reportedly seeking work; the main reason was that there was not enough work (3.01% of workers). This may be treated as an equivalent of disguised unemployment.

Thus, if we take as unemployed the following sum:

(Respondents with Principal status code = 81) + (Respondents with principal status code = 11–51 and not working for at least 6 months) + (Respondents with principal status code = 11–51 and seeking additional work because present work is not enough)

the figure comes to 4.53% of those in labour force. Since the third component is disguised unemployment, we may also drop it. In that case, the unemployment level falls to 2.54%.

seeking work as unemployed. Further, given their non-adult status, it may also be claimed that it is not the children who are seeking work, but their parents/guardians.

Fig. 2 Comparison of levels of unemployment and workers without work for at least six months—by place of residence and gender

So far we have not considered subsidiary status. The NSSO definition uses information on both principal and subsidiary status. So a worker is defined as:

$$(\text{Principal status } 11-81) + (\text{Subsidiary status } 11-51)$$

so that unemployment level is defined as:

$$\frac{(\text{Principal status } 81)}{(\text{Principal status } 11-81) + (\text{Subsidiary status } 11-51)}$$

The definition of unemployment that we propose to use is:

$$\frac{(\text{Principal status } 81) + (\text{Principal status } 11-51 \text{ and without work for at least 6 months})}{(\text{Principal status } 11-81) + (\text{Subsidiary status } 11-51)}$$

In the next section, we shall examine the variation of unemployment—or, as we shall call it, 'being without work'—across economic and demographic covariates, and undertake an econometric analysis to identify its determinants. We will also estimate a model to study what are the factors affecting the duration of being without work.

3 Database and Methodology

The study is based on the employment and unemployment survey conducted in the 68th round of NSSO during July 2011–June 2012. A stratified multi-stage design was adopted for the 66th round survey. The first stage units (FSU) were the 2001 Census villages (Panchayat wards in case of Kerala) in the rural sector and Urban Frame Survey (UFS) blocks in the urban sector. In addition, two non-UFS towns of Leh and Kargil of Jammu & Kashmir were also treated as FSUs in the urban sector. The ultimate stage units (USU) were households in both the sectors. Hamlet groups/sub-blocks constituted the intermediate stage whenever these were formed in the sample FSUs.

The survey was spread over 7,469 villages and 5,268 urban blocks covering 1, 01,724 households consisting of 59,700 in rural areas and 42,024 in urban areas. The survey enumerated 4,56,999 persons, of which 2, 80,763 were in rural areas and 1,76,236 in urban areas. About 69% of the households in India belonged to rural areas and accounted for about 71% of total sample respondents.

The following variables were used in the study:

1. **Place of residence**: This could be either rural or urban.
2. **Age**: Age of respondent.
3. **Education Level**: Information on the education level of respondents was recoded into five categories—illiterate, below primary, secondary completed, higher secondary completed and graduation and above. Under the category 'illiterate' we have included, apart from illiterates, literate persons without formal schooling, and persons educated under adult education schemes like Total Literacy Campaign, etc.
4. **Vocational Training**: Vocational training is defined as any training which prepared an individual for a specific vocation or occupation. Respondents were classified into two groups—those with vocational training and those without vocational training.
5. **Socio-Religious Community**: Respondents were grouped by religion and caste into Hindu-Forward Castes (HFC), Hindu-Scheduled Tribes (HST), Hindu-Scheduled Castes (HSC), Hindu-Other Backward Castes (HOBC), Muslims (Muslim) and other non-Muslim religious minorities (Others).
6. **Household Type**: The household type was decided based on sources of the household's income during the 365 days preceding the date of survey. For this purpose, only the household's income (net income and not gross income) from economic activities was considered; the incomes of servants and paying guests were not taken into account. In rural areas, households were categorised into self-employed in agriculture, self-employed in non-agriculture, regular wage/salary earners, casual labour in agriculture, casual labour in non-agriculture and others.
   In urban areas, the household types are self-employed, regular wage/salary earners, casual labour and others.

7. **Monthly Per capita Household Expenditure**: NSSO collects information on the expenditure of households. This was used to estimate monthly per capita expenditure. We have used a logarithmic transformation of this variable.
8. **Land cultivated**: In rural areas, land cultivated is an important indicator of economic status. Further, land holdings can provide a source of employment.
9. **Membership in the Labour Union**: Union/association meant any registered/recognised body whose membership is open to a section of those engaged in a specific activity or trade and whose main objective is to look after the interests of its members. Working respondents were categorised into three groups—union members, not a union member and worker in the enterprise without any union.
10. **Sub-rounds**: The NSSO survey was undertaken from July 2014–June 2015. The one-year period was divided into four sub-rounds based on each quarter.

4 Unemployment: Levels and Its Determinants

Unemployment level using the NSSO definition and the alternative definition is given in Table 1. The difference in terms of magnitude is not much, but rural unemployment goes up by seven times, while urban unemployment quadruples. These are substantial changes.

Analysis of unemployment levels over sub-rounds (Fig. 3) using the alternative definition shows rural unemployment peaking in July–August (monsoon), followed by a smaller peak in January–March (Kharif). In urban areas, the peak is in the October–December period (monsoon).

Table 1 Estimates of unemployment using the NSSO and the alternative definition—by place of residence and gender (percentage)

Place of residence	Gender	NSSO	Alternative	Increase factor
Rural	Male	0.35	1.65	5
	Female	0.42	5.27	13
	All	0.37	2.75	7
Urban	Male	0.23	0.74	3
	Female	0.48	3.00	6
	All	0.28	1.19	4
All India	Male	0.31	1.38	4
	Female	0.43	4.84	11
	All	0.35	2.34	7

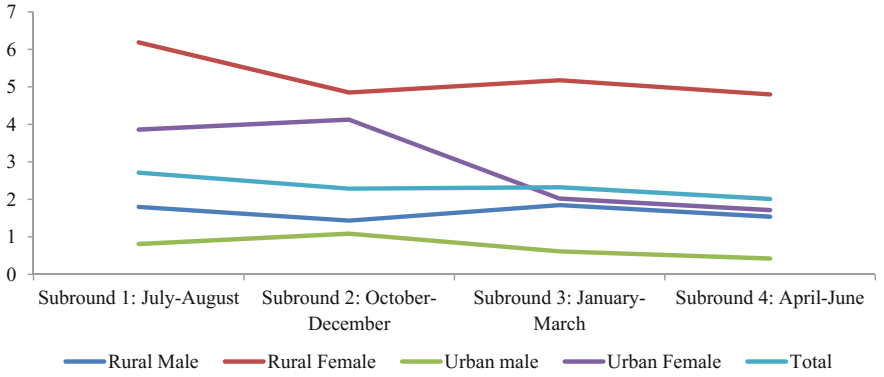


Fig. 3 Unemployment levels by sub-rounds

4.1 Variations in Unemployment by Correlates

Table 2 examines variations in unemployment rates across socio-demographic groups. There is no substantial variation in unemployment across groups. However, in case of certain variables, a relationship with unemployment levels may be seen.

Rural unemployment is highest among HSCs and Muslims. In urban areas, male unemployment is highest among HSCs; if women workers are considered, the unemployment rate is highest among HST workers. The social marginalisation of these groups is thus reinforced by market forces operating through the labour market.

If we consider male workers, unemployment levels are lower in higher expenditure quintiles. Among females, a clear relationship is absent. This may be because males are the main earners in the family and their earnings can compensate for unemployment of female members.

No clear relationship is found between education and unemployment. In case of rural males, unemployment is observed to be higher for those with at least secondary education. In contrast, unemployment levels among urban males and rural females are marginally higher among the more educated respondents. An inverse U shape is observed for urban females. Somewhat surprisingly, unemployment rates are higher among respondents with technical education.

The presence of unions in enterprise reduces the probability of being without work. The unemployment rate is lowest among union members.

Unemployment is high among the residual others category in both rural and urban areas. In rural areas, casual workers (rural males) and wage/salaried category (rural females) also display high rates of unemployment. In urban areas, unemployment is also high among casual workers (both male and female workers).

Table 2 Variations in unemployment rates across socio-demographic groups—percentage

Correlate	Rural male	Rural female	Urban male	Urban female
Hindu FC	1.89	4.30	0.54	2.74
Hindu-OBC	1.42	5.10	0.81	2.94
Hindu-SC	1.88	7.06	1.37	2.34
Hindu-ST	1.04	3.75	0.49	9.46
Muslim	2.07	4.85	0.44	2.92
Others	2.03	6.75	1.01	2.85
Low	1.79	6.91	1.29	3.14
Low medium	1.61	4.98	0.85	3.67
Medium	1.67	4.26	0.63	2.68
High medium	1.68	5.00	0.60	3.31
High	1.43	4.61	0.56	2.26
Illiterate	1.50	5.67	0.73	2.73
Below primary	1.52	4.31	0.89	2.79
Below secondary	1.37	5.15	0.78	4.29
Secondary completed	2.28	4.23	0.55	2.99
Graduation and above	2.25	3.96	0.63	2.22
No technical education	1.64	5.26	0.70	3.02
Technical education	2.64	7.40	1.20	2.80
No union formed	1.80	5.37	0.94	3.48
Not union member	1.29	5.88	0.61	1.99
Union member	1.80	5.37	0.94	3.48
Rural self-employed agriculture	1.56	4.10		
Self-employed non-agriculture	1.45	4.11		
Wage/salaried class	1.09	6.31		
Casual labour in agriculture	2.10	5.63		
Casual labour in non-agriculture	1.69	8.05		
Others	7.24	20.39		
Self-employed			0.71	3.18
Wage/salaried earners			0.69	2.17
Casual labour			0.84	4.31
Others			3.25	9.01

4.2 Econometric Analysis

Table 2 presents the results of a simple bivariate association; it does not control for other variables or study causation. In this section we estimate an econometric model to identify causal factors underlying the phenomenon of being without work. Given the binary nature of the dependent variable, a probit model is estimated. The model structure is as follows:

> Being without work = f (log of per capita monthly expenditure, age and its square, education, technical education, socio-religious identity, land cultivated (in rural areas), union membership)

A problem with this model is that a respondent without work will have low per capita monthly expenditure. The presence of reverse causality between the log of per capita monthly expenditure and being without work implies that the assumption $E(u_i, x_{ji}) = 0$ is violated. Given endogeneity we have used an instrumental variable model, taking the worker/dependent ratio as an instrument.[7] The model has been estimated without sub-round dummies and with sub-round dummies (Table 3). Comparison of results for the two sets of models does not reveal substantial differences.

Unemployment is related to the instrument, worker/dependent ratio in the family. The nature of this relationship, however, varies between rural and urban areas. In rural areas, the probability of unemployment increases with this ratio; a possible cause is that a worker/dependent ratio may result in some workers withdrawing from the labour market for part of the year. For instance, some workers may join the workforce only when seasonal demand for labour from primary activities peaks (during the harvesting season). In urban areas, the relationship between the two is negative.

Studies have reported that aged people are less likely to be unemployed (Love and Torrence 1989; OECD 2017). Although this is confirmed in the study, the relation between the probability of being without work and age is found to be nonlinear, yielding a U-shaped curve. However, in the case of urban females, we find a positive relationship between the two.

In rural areas, HFCs are less likely to be unemployed than the marginalised social groups and Muslims. In urban areas, however, we observe the opposite. This broadly supports findings reported in NSSO reports about variation in unemployment rates across social groups and religious groups (GoI 2006, 2014).

Possessing larger plots of cultivated land reduces the probability of being without work. This is expected as land provides scope for all willing family workers to contribute to production (Grabowski et al. 2013).

The relationship between unemployment and education differs sharply between rural and urban areas. In rural areas, educated respondents are less likely to be without work. In urban areas, on the other, unemployment is higher among the educated respondents. This contradicts Mincer (1993), who had argued that education would

[7]This is defined as number of workers as a proportion of number of children below 15 years and non-working persons aged above 60 years.

Table 3 Results of instrumental probit model—by place of residence and gender

Variable	IV probit model (without sub-rounds)				IV probit model (with sub-round dummies)			
	RM	RF	UM	UF	RM	RF	UM	UF
Worker/dependent ratio (INS)	3.616***	4.468***	−21.410***	−2.925***	2.984***	1.995***	−89.205***	−2.583***
Age	−0.083***	−0.110***	−0.417***	0.003***	−0.087***	−0.070***	−1.300***	0.001***
Age2	0.001***	0.001***	0.007***	0.000***	0.001***	0.001***	0.024***	0.000***
Hindu-Forward Caste (REF. CAT.)								
Hindu-OBC	0.281***	0.587***	−3.076***	−0.426***	0.196***	0.263***	−13.140***	−0.385***
Hindu-SC	0.691***	1.012***	−5.783***	−0.729***	0.565***	0.476***	−24.783***	−0.647***
Hindu-ST	0.876***	1.519***	−6.688***	−0.423***	0.676***	0.591***	−27.353***	−0.308***
Muslim	0.225***	0.230***	−5.600***	−0.850***	0.157***	0.029***	−23.129***	−0.736***
Others	−0.464***	−0.238***	0.311***	−0.012***	−0.380***	−0.110***	0.299***	−0.003
Cultivated land	−0.319***	−0.366***			−0.292***	−0.256***		
Illiterate (REF. CATEGORY)								
Below primary	−0.455***	−0.362***	2.168***	0.321***	−0.417***	−0.232***	8.836***	0.269***
Below secondary	−0.585***	−0.613***	4.308***	0.774***	−0.508***	−0.295***	18.200***	0.697***
Completed secondary	−0.467***	−1.338***	7.340***	1.226***	−0.332***	−0.−0.644***	30.829***	1.097***
Completed graduation	−0.782***	−2.134***	17.044***	2.636***	−0.543***	−0.899***	70.452***	2.323***
No technical education (REF. CAT.)								

(continued)

Table 3 (continued)

Variable	IV probit model (without sub-rounds)				IV probit model (with sub-round dummies)			
	RM	RF	UM	UF	RM	RF	UM	UF
Technical education	0.390***	−0.105***	−2.714***	−0.348***	0.323***	−0.122***	−11.072***	−0.319***
No union (REF. CATEGORY)								
Union present, but not member	−0.843***	−0.616***	0.931***	−0.032***	−0.689***	−0.187***	5.019***	−0.048***
Member of union	−1.620***	−1.212***	1.667***	0.002	−1.466***	−0.859***	8.685***	−0.056***
SR1: July–September (Ref. Cat.)								
SR2: October–December					−0.347***	−0.132***	4.792***	0.227***
SR3: January–March					−0.386***	−0.248***	6.987***	0.030***
SR4: April–June					−0.378***	−0.321***	7.467***	−0.128***
Intercept	−25.214***	−30.078***	160.912***	19.208***	−20.540***	−13.647***	665.658***	16.709***
N	102,805,202	51,219,712	71,941,291	18,777,539	102,805,202	51,219,712	71,941,291	18,777,539
Wald χ^2	1142508.73	290014.55	74866.33	101008.69	1261981.45	385973.35	5696.13	162843.71
χ^2 statistic for exogeneity test	87862.6	3227.39	64960.85	6743.55	84249.7	2846.07	53650.6	2026.13
Instrument: log of worker/dependency ratio								

Note * $p<0.05$; ** $p<0.01$; *** $p<0.001$

Table 4 Results of equidispersion assumption test

Alpha	Coef.	Std. err.	t ratio	Prob>t
RM	1.41	0.02	59.64	0.00
RF	1.25	0.02	72.70	0.00
UM	5.81	0.15	39.03	0.00
UF	2.57	0.05	50.30	0.00

increase avenues of work and reduce chances of being without work. In general, however, studies report that

> higher the education level, the lesser is the likelihood of unemployment in the developed countries. However, in the context of developing countries, such as, Chile, Brazil and Mexico, once the level of education goes up, the unemployment rate also increases. This could be due to a demand or skill mismatch or low absorption capacity of the labour markets in the developing countries vis-à-vis the developed countries (Bairagya 2015: 4).

Technical education, in contrast to general education, reduces the probability of being without work. This is consistent with findings in earlier studies. Technical education facilitates entry from the educational system into the job market; it also increases the productivity of workers and increases chances of innovation (Biavaschi et al. 2013). A study by Cologne Institute for Business Research (Solga et al. 2014) reported that vocational training reduced the rate of unemployment among German youth. This is observed for all groups, except rural males. In rural areas, the limited job market opportunities outside the primary sector—where technical education may be a relevant asset—may result in higher unemployment rates among technically educated respondents.

Unionisation has been known to increase wages. This, however, is supposed to have an adverse long-run impact on employment rates (Olson 1986; Freeman and Medoff 1984). A positive correlation between unionisation and unemployment rate has been reported in the literature for developed countries; this has also been supported by studies in Asian countries (Kim 2005). Our analysis is not consistent with this evidence. Unionisation reduces chances of being without work, particularly if the respondent is a union member.

5 Duration of Unemployment

Another important issue is the duration of being without work. Since the duration is given in months, the dependent variables take the values 1, 2, 3,… 12. A count data model is appropriate in this context. Two competing count data models are Poisson and negative binomial model. The choice between them involves testing the equidispersion assumption that mean and variance are equal. Results of the test (Table 4) indicate that this assumption is not valid, so that the negative binomial model should be used.

Table 5 Results of negative binomial model for duration of being without work

Variable	RM	RF	UM	UF
AGE	−0.068***	−0.032***	−0.090***	−0.021***
AGE^2	0.001***	0.000***	0.001***	0.000***
Hindu-Forward Caste (RC)				
Hindu-OBC	−0.052***	−0.001	0.245***	−0.110***
Hindu-SC	0.037***	0.114***	0.591***	−0.098***
Hindu-ST	−0.098***	0.003*	0.598***	0.647***
Muslim	−0.098***	0.011***	0.180***	0.119***
Others	0.010***	0.096***	0.691***	−0.241***
Cultivated land	−0.163***	−0.161***		
Illiterate (RC)				
Below primary	−0.173***	−0.140***	−0.154***	−0.169***
Below secondary	−0.155***	−0.076***	−0.302***	0.011***
Completed secondary	−0.178***	−0.254***	−0.578***	−0.200***
Completed graduation	−0.145***	−0.336***	−1.026***	−0.577***
No technical education (RC)				
Technical education	−0.085***	−0.257***	−0.167***	−0.386***
No trade union (RC)				
Union present, but not member	−0.433***	0.003*	−0.309***	−0.353***
Member of union	−1.158***	−0.664***	−0.988***	−0.895***
SR1: July–September (Ref. Cat.)				
SR2: October–December	−0.206***	−0.239***	−0.037***	−0.061***
SR3: January–March	0.063***	−0.011***	0.179***	−0.006***
SR4: April–June	−0.262***	−0.222***	−0.194***	−0.206***
Intercept	1.984***	1.353***	1.233***	0.726***
N	78662763	40923194	54684743	14668572
χ^2	1346571.45	422824.29	449340.61	114126.31
Pseudo-R^2	0.05	0.02	0.08	0.05

When estimating the negative binomial model we have taken all variables used in the probit model, barring log of per capita monthly expenditure. Results are given in Table 5.

Results indicate that young workers are likely to be without work for longer periods. This, however, is not a linear relation; rather, the positive and significant coefficient of the square of age indicates an inverse U shape. HOBCs, HSTs and Muslims are without work for a shorter duration, while HSCs are out of work for longer periods compared to HFCs (rural males). Among rural females, marginalised groups (HSCs, HSTs and Muslims) are without work for longer periods. In urban areas, HFCs are without work for shorter periods (male respondents). Among urban females, HSTs and Muslims are unemployed for longer periods, while HOBCs and HSCs are without work for shorter periods, compared to HFCs. Illiterates are without work for longer periods; so are respondents without technical education. Respondents are likely to be without work for long periods if their last enterprise does not have any union, or if they are not union members.

6 Conclusion

Despite the economic and social importance of unemployment, political considerations have created a tendency to under-report levels of unemployment in India. This is achieved by adopting a narrow definition of unemployment. In this study, we argue that an alternative definition of unemployment, based on the same surveys used to report unemployment, can produce somewhat more realistic estimates of unemployment levels.

The definition of unemployment proposed by us is to use the information on duration over which the respondent is without work in the year preceding the survey. We argue that policymakers and researchers have failed to utilise the potential of this information. This information can not only generate more realistic levels of unemployment, but also provide information on the duration of unemployment. Both are important parameters with substantial implications for welfare levels of the population. Further, this information is provided in earlier rounds also; hence, it may be used to understand the temporal nature of the labour market and its ability to absorb the increasing number of workers. This exercise has not been addressed in the current study but is one that may be addressed in the future.

References

Bairagya, I. (2015). *Socio-economic determinants of educated unemployment in India*. Working Paper No 343, Bengaluru: Institute of Social and Economic Change.

Biavaschi, C., Eichhorst, W., Giulietti, C., Kendzia, M. J., Muravyev, A., Pieters, J., et al. (2013). *Youth unemployment and vocational training*. The World Bank: Washington D.C.

Brand, J. E. (2015). The far-reaching impact of job loss and unemployment. *Annual Review of Sociology, 41,* 359–375.

Freeman, R., & Medoff, J. (1984). *What Do Unions Do?*. New York: Basic Books.

Gangl, M. (2006). Scar effects of unemployment: An assessment of institutional complementarities. *American Sociological Review, 71*(6), 986–1013.
GoI. (2012). Report on employment and unemployment survey 2011–12. Chandigarh: Ministry of Labour, Labour Bureau.
Government of India. (2001). *Concepts and definitions used in NSSO*. New Delhi: National Sample Survey Office.
Government of India. (2006). *Employment and unemployment situation in India: NSS 61st round (July 2004–June 2005)*. Report No. 515(61/10/1), New Delhi: National Sample Survey Office.
Government of India. (2014). *Employment and unemployment situation in India: NSS 68th round (July 2011–June 2012)*. Report No. 554(68/10/1), New Delhi: National Sample Survey Office.
Grabowski, R., Self, S., & Shields, W. (2013). *Economic development: A regional, institutional, and historical approach*. New York: Routledge.
Gregory, M., & Jukes, R. (2001). Unemployment and subsequent earnings: Estimating scarring among British men 1984-94. *Economic Journal, 111*(475), F607–F625.
Kim, D. K. (2005). Unionization, unemployment, and growth in Korea: A cointegration approach. *Atlantic Economic Journal, 33*(2), 225–233.
Love, D. O., & Torrence, W. D. (1989). The Impact of Worker Age on Unemployment and Earnings After Plant Closings. *Journal of Gerontology, 44*(5), S190–S195.
Marmot, S. M., Bloomer, E., & Goldblatt, P. (2013). The role of social determinants in the tackling health objectives in a context of economic crisis. *Public Health Reviews*, *35*(1), 1–24. Mincer, J. (1993). Job training: costs, returns, and wage profiles. In Mincer J. (Ed.). *Studies in human capital* (263–281). Aldershot: Edward Elgar.
Nordhaus, W. (1975). The political business cycle. *Review of Economic Studies, 40*(2), 169–190.
OECD. (2017). *Ageing and employment policies*. http://www.oecd.org/employment/ageingandemploymentpolicies.htm. Retrieved August 12, 2017.
Olson, M. (1986). Maintaining a healthy business climate: A broader perspective on the rates of economic growth and of unemployment in the southern and southwestern states. In *Energy and the Southwest Economy: Proceedings of the 1985 Conference on Energy and the Southwest Economy*. Dallas: Federal Reserve Bank of Dallas.
Ruhm, C. J. (1991). Are workers permanently scarred by job displacement? *American Economic Review, 81*(1), 319–324.
Sen, A. K. (1975). *Employment technology and development*. Oxford: Clarendon Press.
Solga, H., Protsch, P., Ebner, C., & Brzinsky-Fay, C. (2014). *The German vocational education and training system: Its institutional configuration, strengths, and Challenges*. Discussion Paper SP I 2014–502, Bonn: University of Cologne and Federal Institute for Vocational Education and Training.
Stewart, F., & Streeten, P. (1971). Conflicts between output and employment objectives in developing countries. *Oxford Economic Papers, 23*(2), 145–168.
Tøge, A. G. (2016). Health effects of unemployment in Europe during the great recession: The impact of unemployment generosity. *International Journal of Health Services, 46*(4), 614–641.

Part III
Sustainability and Corporate Governance

Sustainable Eco-Management: Participatory Mechanisms and Institutions

Sarmila Banerjee

Participatory approach in managing development projects and programs in poor countries has emerged in response to global demands for greater individual and societal control over the activities of state and private agencies, and especially to the overt failures of traditional 'top-down' management systems in LDCs (Brett 2003). Beneficiary participation is a special form of decentralized development management where the development projects are locally implementable and the micromanagement is coordinated with the macro-priorities to ensure long-term sustainability. Here, not only do the local authorities take decisions at the local level but there is a harmonious interaction among the provider, regulator and the service recipient (Chopra and Kadekodi 1991; Paul 1989). In fact, it is a cooperative management approach where all stakeholders would take up an active role and enjoy equal status in the process of decision making and execution. The programs are motivated mostly by livelihood approach with high values attached to an enhanced access to social capital. Drawing examples from Canadian inland small-scale fisheries, it has been shown by Berkes (2003) that instead of a top-down experts-know-best type control-oriented management strategy, a participatory approach can guarantee better sustainable livelihood where social and ecological combination of governance practices through coping, learning and adaptation is capable of delivering more resilient outcomes.

However, it has been shown by many social scientists that though the participatory approach is conceptually more democratic, its success potential is highly dependent on the local conditions. Experience shows that co-management regimes can set into motion new conflicts or cause old ones to escalate. Instead of contributing to local empowerment, such arrangements may further marginalize communities and resource users (Castro and Nielsen 2001). Sarkar and Sinha (2015) proposed an exploratory framework to study development management from the perspective of strategic interventions where the specific requirements for the success of the

S. Banerjee (✉)
University of Calcutta, Kolkata, India
e-mail: sarmila_b@yahoo.co.in

P. Ray et al. (eds.), *Economics, Management and Sustainability*,
https://doi.org/10.1007/978-981-13-1894-8_9

participatory approach have been nicely schematized. The participatory approach commonly fails in contexts where local conditions make cooperative and collective action very difficult, or where it is manipulated by implementing agencies to justify their own actions and/or poor performance. In practice, the result may not be power sharing but rather a strengthening of the state's control over resource policy, management and allocation. Instead of contributing to local empowerment, such arrangements may further marginalize communities and resource users and make the program non-viable.

This chapter draws on a number of studies conducted by the present author and her co-researchers between 2007 and 2014 on different rural, peri-urban and urban pockets of the state of West Bengal, India, focusing on ecologically sustainable management of natural resource and environment like social forestry, wetland fisheries and municipal solid waste disposal in a co-management framework with active beneficiary participation.[1] Since the contexts are distinct, there was variation in the notion of sustainable management itself and in the composition of the stakeholder groups. The design of instruments was innovative at times, involving intensive use of local wisdom and specificities with active cooperation between local people and the regulatory authority, but sometimes the planning and management processes faced bottlenecks due to conflict of interest among the members of the stakeholder group leading to insurmountable stalemates. All these studies were conducted to explore the context-specificity of the success probability of co-management practices in different situations. We documented, particularly, two different situations where otherwise suitable projects for participatory resource management failed to attain the intended result due to some peculiarities related to the cases related to the presence of some built-in contradictions either in the composition of the stakeholder group leading to a deviation in the equal participation norms or in the regulatory setup comprised of multiple authorities with inherent jurisdictional conflicts.

In most cases, managing development involves multiple objectives and long-run sustainability is only one aspect, though a very important dimension. When the stakeholders are pursuing these objectives with similar (relative) priorities, the situation is more manageable. However, in case of different priority rankings, defining a feasible and mutually agreeable core itself becomes a problem. There the voice of some particular group starts dominating the others and the entire philosophy of participatory co-management gets defeated. To illustrate this situation, we have taken up a study on the joint forest management practices in Birbhum Forest Division of West Bengal, India, in 2012.

Another problem creating hindrance in smooth functioning of participatory development management practices was identified as the overlap of administrative jurisdictions of different regulatory authorities (Gilbert 2006). For example, the city of Kolkata is facing rapid east-bound expansion over time leading to conversion of

[1] The study on social forestry was undertaken jointly with Jayita Bit, St. Xavier's College, Kolkata; the study on wetland fisheries was taken up with Debanjana Dey of Adamas University, Kolkata, and the collaborators for the study on municipal solid waste management were Prasenjit Sarkhel of Kalyani University, Nadia, and Somdutta Banerjee of the Institute of Economic Growth, New Delhi.

wetlands into urban settlement. These wetlands generally play a very important role in providing the city an option of natural sewage treatment and absorbing its excess rainwater run-off. Besides that, the sheer existence of this wetland provides livelihood support to a large number of fishermen and helps to maintain the ecological health of the entire southern Bengal through a series of wise use practices adopted by the local residents by using their traditional knowledge. So, the direct benefits of land conversion need to be balanced against the huge loss of eco-system services involved in this decision. The problem with that type of assessment mostly emanates from the non-market nature of this loss where the price of lost ecological balance is really difficult to measure as it would be realized only in the long-run when the service would no longer be available. So, the land-use pattern in this area needs to be controlled judiciously and a number of administrative authorities are working toward that end. It has been observed through a series of field surveys conducted in the East Kolkata Wetlands over the period 2012–2014 that because of the active involvement of Kolkata Municipal Corporation (KMC), Kolkata Metropolitan Development Authority (KMDA), West Bengal Housing Infrastructure Development Corporation (WBHIDCO), NGOs, Fishermen cooperatives, local residents, eco-tourism resort owners, real estate developers, and finally wetland institutes like the Institute of Wetland Management and Ecological Design (IWMED), East Kolkata Wetland Management Authority (EKWMA) with definite hierarchy in their administrative power, the grass root level organizations have very little say in the ultimate decisions leading to an unintended alienation of the direct users and therefore, direct victims/beneficiaries from the planning and management process.

A third situation attracting our attention is related to the provision of an urban amenity service like solid waste management where beneficiary participation unfolded some interesting dimensions related to their perception about service quality that was otherwise unknown to the provider and which played a strategic role in enhancing the success probability of such co-managed projects. On the onset of the adoption of the Municipal Solid Waste Management and Handling Rule (2000) in urban West Bengal intending a separation of biodegradable and non-biodegradable waste at the household level and encouraging public–private partnership in solid waste management, a primary survey was conducted in the Bally Municipality of West Bengal in 2006. After partial implementation of the Rule, a repeat survey was carried out on the same set of households in the year 2011. In both cases, the household's willingness to pay toward the project cost was estimated by applying a contingent valuation method. Though most of the households admitted an improvement in the environmental condition of the locality after the implementation of the Rule, there was an overall decline in the willingness to pay for this service. This apparently contradictory finding indicates the role of service quality in making the project a success. In the pre-project days, the WTP quotes depend on the notional service quality expected to be provided and in the post-project days the WTP was revised in terms of the actual service quality received. The decline in the WTP is an indicator of this dissatisfaction on the part of the benefit recipients. The following paragraphs are elaborating on these studies.

1 Study I: Joint Forest Management[2]

Forest resource represents a partly open access common property resource serving different ecological, social and economic purposes in the immediate-run as well as distant future involving multiple stakeholders with variedly different objectives and preferences regarding the efficient management of the resource. These stakeholders pursue different and sometimes mutually conflicting objectives like sustainable forest development, expansion of livelihood support, enhancement of economic opportunities and have different expectations from the efficient management of the forest resource. In appreciation of this inherent tension, a participatory approach to forest management has gained wide acceptance all over the globe. It started in India since the inception of Joint Forest Management (JFM) program through National Forest Policy in 1988. However, low levels of involvement, unwillingness among stakeholders to participate in forest planning processes and general dissatisfaction are a commonly observed picture over the past few years.

A successful forest management plan is essentially a multi-objective-programming involving multiple stakeholders where the preference ranking among different objectives varies across stakeholders and the existence of a core in terms of mutually agreed feasible solution may be difficult to ensure (Kangas and Kangas 2005). Multi-criteria decision making technique is considered a suitable framework for program-evaluation in this context as it has the potential to incorporate both conflict and multidimensionality within a single integrated structure. To assess the suitability of participatory approach for social forestry, a study was taken up in 2012 on the functioning of the Joint Forest Management scheme in the Birbhum district of West Bengal, India.

Birbhum Forest Division: The field survey was conducted during September–November, 2012 in the district of Birbhum. From the administrative perspective, this district is divided into 7 forest ranges (Bolpur, Md. Bazar, Rajnagar, Rampurhat, Sainthia, Dubrajpur and Suri) containing 19 beat offices. Among them, 6 beats under 5 ranges were randomly selected for the present study to cover 10 locations with presence of active Forest Protection Committees (FPC) and representing variations in topographical properties, cultural heritage as well as nature of forest biodiversity. Ranges with limited forest cover were excluded from the sample. The prospects and problems of participatory forest management were assessed in terms of response to three basic queries: (i) whether the forest-dependent people have benefitted from the present policies; (ii) whether the new form of plantations are enough to meet the needs of forest-dependent people; (iii) whether all the stakeholders take active interest in forest and its biodiversity conservation.

In all, ten FPCs have been surveyed maintaining variations with respect to the percentage of tribal members, proximity to township, quality of FPC performance, etc. From each of the locations, ten households were randomly selected having at least one family member in FPC and to isolate the impact of this FPC participation on

[2]The discussion presented in this section draws heavily from Bit and Banerjee (2015), re-used here with due permissions.

forest-dependent livelihood, conservation of traditional knowledge and sustainable forest management, similar information was collected from ten households having no FPC member. Information pertaining to personal details of household members, condition of the nearby forest area, general awareness regarding the role of forest in supporting livelihood, importance of conserving traditional knowledge, etc., was collected from a total of 204 households. The questionnaire for Forest Department and Biodiversity Board solicited information about the objectives, concern and process of the present forest management practice. Similar questionnaire was designed for the FPC groups. Gram Panchayat members and eco-tourism developers were interviewed about the roles they play in forest conservation, and there were open group discussions with all groups. Finally, the group specific perception of social, economic and ecological importance of forest has been attempted to be assessed through ranking of preferences over wide range of qualitative responses.

1.1 Stakeholder Analysis: A MCDA

Forest resource planning is about multi-objective programming. It addresses the issue of optimizing several objectives perceived with different relative importance, subject to a set of constraints. Given any conflict among the set of objectives, which is rather commonplace, not all of them can be simultaneously optimized. Instead of searching for a nonexistent optimizing solution, it is considered better to search for a consistent solution which would be efficient in the sense of Pareto-optimality where no other feasible solution can improve one objective without degrading at least one other (Diaz-Balterio and Romero 2008). In making plans and managing natural resources, multi-criteria decision analysis helps to consolidate the multiple views and knowledge of stakeholders to support decision making in complex environments. Criteria and indicator provide a common framework to describe, conceptualize, organize and interpret information related to sustainable forest management. The hierarchical structure of defining indicators allows a complex problem to be broken down into manageable elements that can lend themselves to formal analysis (Mendoza and Prabhu 2005; Mendoza and Martins 2006; Khadka and Vacik 2012). In this context, the methodology of analytical hierarchy process (AHP) is a robust, ratio-scaled MCA method for analyzing complex decisions with multiple attributes (Saaty 1980).

Sustainable forest management usually involves the use of criteria and indicators (C & I) allowing the monitoring, reporting and assessment of management activities at national, regional and forest management unit levels. From the official documents of the Forest Department in India, three important objectives or criteria of forest management can be identified as (a) environmental concern like conservation of forest, (b) social concern like protection of livelihood and (c) economic concern like extraction of marketable benefits. The strategic instruments available for forest conservation are (i) enhancement of canopy cover, (ii) protection of biodiversity and (iii) conservation of soil and water, which can be taken as three indicators. Similarly, for livelihood support the most important indicators are (i) collection of non-timber-

forest-products (NTFP), (ii) conservation of medicinal plants (via transmission of the flow of local knowledge and traditional wisdom) and (iii) expansion and protection of forest-dependent livelihood options. Finally, to enhance economic benefits interventions can be designed (i) to raise revenue from timber sales, (ii) increase the supply of logs and timbers and (iii) promote the development of eco-tourism, and these can be taken as indicators of economic usages of forest resources. So, in all, there are three criteria and nine indicators of forest development which are considered to be of strategic importance to achieve the goal of sustainable joint forest management.

In this situation, six different stakeholders can be identified as (I) the Forest Department (FD) at the national level, (II) the Gram Panchayat (GP) at the block level, (III) the members of the Forest Protection Committee (FPC) at the local level, (IV) the households where there is no member of the FPC (non-FPC), (V) the Eco-tourism Developers/Resort Owners (ETD) and (VI) the Timber Merchants (TM). In our field survey, the different types of stakeholders are asked to rank their preference for each indicator of forest development in a 5-point rating scale of [1–5], where 5 indicates the most important and 1 indicates the least importance. We have interviewed 10 officials of the forest department, 6 office bearers of the GP, 73 members of the FPC, 52 households with no FPC member, 3 resort owners and 3 Timber merchants. Depending on all these 147 responses and by combining a group decision making (GDM) approach with that of multi-objective programming (MOP), a multi-criteria decision model (MCDM) has been developed to assess the consistency and compatibility of the stakeholders' attitudes necessary to make the joint forest management program sustainable. Our specific objectives are twofold:

(a) To assess the relative importance of a proposed set of C&I with respect to sustainable forest management by elicitation of stakeholders preferences;
(b) To evaluate perceptions of the overall performance of community forest management strategies by the use of AHP.

Criteria–Indicator Analysis: The stakeholders were asked a qualitative assessment of alternative indicators under each criterion using a five-point scale, and the average response in each case has been estimated by taking the geometric mean (GM).[3] If P is the number of *stakeholder types* indexed by k, n_k is the number of *observations* in group k, M is the number of *criterion* indexed by i, and N is the number of *indicators* under criterion i, indexed by j, then:

$GMI_{kij} = \left[\prod_{n_k} I_{kij}\right]^{1/n_k}$, which represents the average ranking for the jth indicator of criterion i for stakeholder type k. Combine these $GMI_{kij}, \forall j$ by taking the GM for criterion i as

[3] In order to allow a synthesis of the individual group priorities with the analytical hierarchy process at the subsequent stages, the judgments have to be combined in a manner so that the reciprocal of the synthesized judgments is equal to the synthesis of the reciprocals of these judgments (Saaty 2008). Among the commonly available measures of average, geometric mean satisfies this requirement (Forman and Peniwati 1998).

Table 1 Ranking-based preference ordering for different criteria and indicators

Goal: sustainable forest management							
Stakeholders → / Criteria/indicators ↓	FD	GP	FPC member	Non-member	ETD	TM	All
C1: Forest conservation	**4.45**	**4.14**	**3.91**	**4.23**	**5.00**	**3.76**	**4.08**
I-1.1: Canopy cover	4.89	4.47	4.89	4.64	5.00	5.00	4.79
I-1.2: Biodiversity	4.05	4.82	3.81	4.36	5.00	4.64	4.11
I-1.3: Soil and water	4.44	3.30	3.20	3.74	5.00	2.29	3.46
C2: Livelihood support	**3.55**	**3.93**	**3.56**	**3.94**	**4.50**	**3.40**	**3.62**
I-2.1: Collection of NTFP	3.40	3.84	4.59	4.48	4.22	3.42	4.39
I-2.2: Conservation of medicinal plants	3.88	4.31	3.08	3.81	5.00	3.68	3.30
I-2.3: Enhancement of livelihood options	3.40	3.66	3.20	3.59	4.31	3.11	3.27
C3: Economic opportunity	**3.07**	**3.20**	**2.82**	**2.32**	**2.96**	**4.50**	**2.79**
I-3.1: Revenue from timber	3.15	3.20	3.70	3.42	2.29	4.64	3.63
I-3.2: Supply of log & timber	3.19	3.14	2.32	2.13	3.30	5.00	2.55
I-3.3: Ecotourism development	2.89	3.26	2.62	1.73	3.42	3.91	2.35

Source Bit and Banerjee (2015)

$GMC_{ki} = \left[\prod_j^N GMI_{kji}\right]^{1/N}$, to get the average ranking of criterion i for the stakeholder group k; and finally, combine these GMC_{ki}, $\forall k$ by taking the GM across k as

$GMC_i = \left[\prod_k^P GMC_{ki}\right]^{1/P}$, to get the average *overall* ranking of criterion i across all stakeholder groups. The ranking-based preference ordering is reported in Table 1.

Except for the timber merchants, for all other stakeholders, forest conservation enjoys higher ranking compared to the livelihood support and livelihood support enjoys higher ranking than economic benefit. However, the magnitude of average differences is not uniform across both indicators and stakeholders. This heterogeneity of average response will influence the relative preference ordering of each group across alternative choice criteria. That would be evident from the matrix developed on the basis of pair-wise comparison (PWC) to obtain the priority ranking.

1.2 Analytical Hierarchy Process

This is a kind of value function approach to assign weights to different criteria by defining suitable priority functions, i.e., different criteria and indicators for different stakeholders are arranged in a hierarchical structure. Preference for any particular criterion (indicator) of any particular stakeholder has already been elicited in a 5-point preference scale with 1 representing the worst and 5 representing the best. We have to derive a relative scale by using the judgments expressed in terms of this standard scale by carrying out pair-wise comparison of different criteria (indicators). Relative priorities of criteria with respect to the overall goal and those with respect to alternative indicators are to be calculated from the corresponding pair-wise matrices.

As an illustration of the methodology of deriving the relative scale, let us consider three forest management-related criteria C_1, C_2 and C_3. Define a_{ij} as the value of the difference in the average rating of C_i and C_j, i.e., the relative importance of C_i over C_j to stakeholder k in achieving the goal. The difference will vary between [0, 4], and it can be mapped into a 9-point scale starting from equal importance, moderately preferred, strongly preferred, very strongly preferred and extremely preferred. If it is 0, then both the objectives are equally important and the decision-maker is indifferent between the two alternatives. If the difference is 1, then it is moderately preferred and a numerical value of 3 will be assigned to represent this relative priority. Similarly for difference of 2, the assigned value would be 5, for 3 it would be 7, and for 4 it would be 9. The in-between values like 2, 4, 6 and 8 would be considered as intermediate values between two adjacent judgments when compromise is needed.

Three important consistency conditions that the assigned values need to satisfy are as follows:

(a) A comparison of criterion C_i with itself is equally important, $\rightarrow a_{ii} = 1$.
(b) Since a_{ij} represents the relative priority of i over j, a_{ji} will represent just the reverse and will be represented by the reciprocal of the original value.

$$\rightarrow (C_i - C_j) = a_{ij} \& (C_j - C_i) = 1/a_{ij}.$$

These two conditions confirm reflexivity axiom of preferences.

(c) The relative scale should satisfy the transitivity property in a cardinal way, i.e., if C_i is 3 times preferred to C_j and C_j is 3 times preferred to C_k, then C_i is (3×3) 9 times preferred to C_k; $\rightarrow a_{ij}.a_{jk} = a_{ik}$. Transitivity along with reflexivity ensures that the consumers are rational and consistent in their preferences.

Let w_i denote the true value of selecting a criterion i out of M alternatives. Then, the relative weight to be assigned to C_i is $a_{ij} = \frac{w_i}{w_j}$. So, $a_{ij}\frac{w_j}{w_i} = 1 \& \sum_j a_{ij}\frac{w_j}{w_i} = M$; or, $\sum_j a_{ij} w_j = M w_i; j = 1, 2, \ldots, M$; which is the jth row of the system of equations $Aw = Mw$;

$$Aw = \begin{bmatrix} 1 & a_{12} \ldots & a_{1M} \\ a_{21} & 1 \ldots & a_{2M} \\ . & \ldots\ldots & \ldots \\ a_{M1} & \ldots\ldots & a_{MM} \end{bmatrix} \begin{bmatrix} w_1 \\ . \\ . \\ w_M \end{bmatrix} = \begin{bmatrix} \sum_{j=1}^{M} a_{1j} w_j \\ . \\ . \\ \sum_{j=1}^{M} a_{Mj} w_j \end{bmatrix} = \begin{bmatrix} Mw_1 \\ . \\ . \\ Mw_M \end{bmatrix} = Mw$$

Since A is an $(M \times M)$ matrix, w is an $(M \times 1)$ vector, and M is a scalar, the system represents a typical characteristic equation with M as the (maximum) eigen value and the priority weights for different criteria given by the components of the associated eigen vector w. This condition is very useful in verifying the consistency of the pair-wise comparison matrix A when the true value w_i's are not known but estimated on the basis of the elicited stated preferences.

For this constructed matrix A^*, calculate the maximum eigen value λ^* and consider the difference between λ^* and M. Define Consistency Index $CI = (\lambda^* - M)/(M - 1)$ and allow for some random deviations for a given M. Denote this Random Index by $RI(M)$. Combining CI and RI, the consistency ratio CR can be obtained as $CR = \frac{CI}{RI}$.. A value of $CR \leq 0.1$ is acceptable.

1.3 Discussion

Table 2 reports the results from an application of AHP in our context of sustainable management of forest in the district of Birbhum, West Bengal, India. It is apparent from the table that all calculations of weights are consistent and except for the timber merchants forest conservation is the most important purpose of forest management to all other stakeholders followed by Livelihood Protection and Economic Opportunities. However, the pair-wise comparison techniques are applied to the indicators of forest conservation and it turned out that the protection and expansion of canopy cover is the most important concern for the Forest Department, FPC members, non-members as well as the timber merchants. Here, Gram Panchayats and the resort owners (eco-tourism developers) are revealing slightly different behavior; while the former considers biodiversity conservation as a very important component of forest conservation, the latter assigns equal weight on all the constituent indicators. In case of livelihood protection, collection of non-timber-forest-products (NTFP) enjoys very high weight from the household sector, both FPC members and non-members.

The remaining groups of stakeholders are more inclined to protect the medicinal plants and thereby the flow of traditional knowledge. Finally, for economic benefits both FPC members and non-member households value revenue earning from timber the most. In the overall rating, the indicator weight of this component is 50% of the total. For timber merchants, this economic benefit is the most important criterion

Table 2 Priority weights generated by AHP

Stakeholders -> /Criteria, indicators	FD	GP	FPC member	Non-member	ETD	TM	All
Forest conservation	**0.62**	**0.43**	**0.52**	**0.56**	**0.57**	**0.30**	**0.54**
Canopy cover	0.54	0.32	0.62	0.54	0.33	0.56	0.56
Biodiversity	0.16	0.56	0.24	0.30	0.33	0.36	0.32
Soil and water	0.30	0.12	0.14	0.16	0.33	0.08	0.12
Livelihood protection	**0.24**	**0.43**	**0.33**	**0.35**	**0.33**	**0.16**	**0.30**
Collection of NTFP	0.25	0.25	0.67	0.55	0.21	0.31	0.40
Protection of medicinal plants	0.50	0.50	0.17	0.24	0.55	0.49	0.40
Expansion of forest-dependent livelihood	0.25	0.25	0.17	0.21	0.24	0.20	0.20
Economic opportunity	**0.14**	**0.14**	**0.14**	**0.09**	**0.10**	**0.54**	**0.16**
Revenue from timber	0.40	0.33	0.62	0.66	0.14	0.30	0.50
Supply of log & timber	0.40	0.33	0.14	0.21	0.43	0.54	0.25
Ecotourism development	0.20	0.33	0.24	0.13	0.43	0.16	0.25
$\boldsymbol{\lambda_{max}}$	**3.02**	**3.01**	**3.07**	**3.05**	**3.02**	**3.01**	**3.02**
$\mathbf{CI} = \frac{(\lambda \max - M)}{(M-1)}$	0.01	0.01	0.03	0.03	0.01	0.01	0.01
$\mathbf{RI(M) = RI(3) = 0.58}$ **&** $\mathbf{CR} = \frac{CR}{RI}$	**0.02**	**0.01**	**0.06**	**0.05**	**0.02**	**0.01**	**0.01**

Source Bit and Banerjee (2015)

where there concern is more focused on regularity of the supply of logs and woods from the forest.

2 Study II: Wetland Fisheries[4]

The natural gradient of the city of Kolkata is toward southeast and, therefore, the 12,500-hectare-wide wetlands in the Eastern fringe of the city, popularly known as the East Kolkata Wetlands (EKW), are providing the city a unique opportunity for natural treatment of its wastewater and flushing out of its excess rain water. Through

[4]The discussion presented in this section draws heavily from Dey and Banerjee (2013a, b, 2015, 2016a, b, 2017) and Banerjee and Dey (2017); Re-Used Here with Due Permissions.

the algae-bacteria symbiosis process, the wastewater is getting converted into clean water, rich with nutrients congenial for fish production. The freshwater aquaculture and garbage farming help to maintain the low-cost supply chain of fish, vegetables and other food items to the city and keep its Cost-of-Living Index (CLI) the lowest among all metro cities of India. The sewage-fed fisheries are operating in this area for more than a century, and the wise use practices emerged through close interaction with nature created a number of vocations and livelihood practices which are not only inter-temporally viable but also socially and economically sustainable. That is why the EKW has been declared as a Ramsar site in 2002 and it got the recognition as a wetland of international importance. The sustenance of the city of Kolkata and that of the EKW are inter-connected through a circular flow of mutual benefits.

2.1 Strategic Importance of Sewage-Fed Fisheries

When the sewage water arrives in the pond network through the inlet channels, it is kept standing in the sun, which results in biodegradation of the wastes through an algae-bacteria symbiosis. In fact, retention of wastewater in the ponds before the initial stocking of fish for a considerable period allows bacteria to work upon the organic waste. The algae that thrive in these shallow ponds under the ample sunshine support the growth of these planktons. However, the overgrowth of planktons becomes a problem for aquatic environment as they lead to algal bloom. At this critical stage, the integration of wastewater treatment and sewage-fed fisheries creates a win–win situation. Fish plays an important role by consuming this extra plankton as food under a very balanced practice of aquaculture. From their traditional knowledge gathered through years of experience, the local fishermen know exactly how to excavate the ponds to the correct depth, clean the water by spraying kerosene, lime, etc., mix the right quantity of sewage, allow optimal time for conversion of the waste into fish feed, the right time to add spawns, protect the embankments through water hyacinths, and finally, through their livelihood practices support the wastewater treatment and support their own living at one go (Dey and Banerjee 2013a). Waste recycling in EKW involves four principal resource recovery practices, viz. sewage-fed fisheries, paddy cultivation by utilizing fish pond effluents, organic waste-based farming of vegetables and freshwater aquaculture. Thus, the cumulative social and economic gain from such sustainable management in an eco-friendly way is crucially contingent on the livelihood dependence of the local people on these traditional vocations, especially fisheries (Yan et al. 1998).

However, since the development of Salt Lake Township and the construction of Eastern Metropolitan Bypass, the connectivity of EKW with the main city has improved and the pressure of urbanization is leading to conversion of some of the water bodies into urban settlements. In fact, the land-use pattern has changed significantly over the last two decade, especially in areas with close proximity to the city of Kolkata. Three broad types of changes can be easily identified: (i) from small water body to settlement, (ii) from agricultural area to settlement and finally and

(iii) from open spaces to settlement areas. Vulnerability is highest in case of mouzas with high population density and small *bheris.* This practice is disturbing the age-old eco-balance and the eco-system-based livelihood in the area and unleashing a dynamically unstable spiral where approach roads are obstructing wastewater canals through newly developed culverts, making the sewage water pisciculture less profitable, creating a number of new vocations related to the process of land speculation, land transfer and urban living. Consequently, the speed of land conversion increases creating big environmental threat for both the city of Kolkata and EKW. If the fisheries lose their pivotal importance in the local livelihood pattern, then the continuity of this ecologically subsidized sewage treatment facility (Ghosh 2005) may no longer be available to the city of Kolkata. So, the influence of urban invasion from the growing city of Kolkata on the vocational choice of the local people could be of serious concern for continuity of this option of natural sewage treatment facility for the city dwellers.

As a protectionist measure, any further change in the pattern of land use in the core area has been legally prohibited. Conserving the landscape of the wetlands does not necessarily ensure the sustenance of its eco-characteristics until and unless people can retain their dependence on the system for their livelihood practices. In fact, the delicate chain of interdependence, once disturbed, will affect the whole system through a multiplier sequence and the initial eco-balance will be almost impossible to regain. In spite of legal strictures, a significant change in the pattern of land use has been observed in the buffer area which has expanded the set of vocational options to the residents of the EKW and encouraged them to go for vocational switching.

2.2 Vocational Transition: Evidence from Field

Two factors are simultaneously influencing the vocational choice, which is a direct offshoot of this change in land-use pattern: The push factor is creating incentive for the local people to move away from the traditional vocations as it is becoming less remunerative, and the pull factor is attracting them toward alternative modern vocational options created in the newly urbanized neighborhood, which are relatively more intensive in modern skills but generating higher return as well. To investigate the real situation, three vocation-related surveys have been conducted in EKW between 2012 and 2014. The first survey was conducted in 2012, only on the randomly picked up workers spread over 17 mouzas, out of which 10 experienced significant change in land-use pattern but not the remaining 7. Total sample size was 325, out of which 42% reported engagement in traditional vocations and the remaining 58% in modern vocations. As expected, within the traditional vocation, fisheries dominate (67% of traditional type) and that again concentrate mostly in the low-change areas. However, what appears to be the most striking observation is that irrespective of the extent of land-use change, for all selected mouzas more than 50% of the respondents were engaged in their present vocation for a period less than 10 years (Dey and Banerjee 2013b). So, either they were new entrants in the labor market or they

had switched their vocation. This is indicative of the fact that the pull effect that is likely to be uniform for all mouzas is dominating the push effect, which should be more prominent in the high-change areas only. To verify that hypothesis with greater analytical rigor focus was laid on the low-change and no-change mouzas and two subsequent surveys had been undertaken to study the socioeconomic and demographic factors influencing the likelihood of vocational switching.

One problem encountered in conducting a survey on vocational engagement and its pattern of change is the incompleteness of the responses. During the field visit, it was observed that generally people are engaged in multiple vocations and in reporting their occupation they are mentioning their major engagement only. Instead, if the respondent could be asked to choose the types of work he/she carries out from an exhaustive list of possible engagements, then the response bias could have been controlled better. In a tradition-bound society, especially for unpaid work provided by the members of the family, it is difficult to distinguish between work and economic activities. Most of the unpaid economic activities are carried out as routine chores, and they themselves cannot consciously recognize all these activities as 'work.' Households have been identified from the low-change and no-change mouzas[5] with literate members in 2013, and all adult members have been requested to fill in time diary for 10 consecutive days during the peak season (August–September) of fish cultivation. This time diary survey helped to generate an exhaustive list of economic activities carried out by the residents of EKW. By mapping those activities on vocational scheme, an exhaustive vocation listing has been generated where vocations have been classified as TRADITIONAL, containing(a) fisheries and related [F], (b) agriculture and related horticulture [A], (c) duck rearing, poultry and animal husbandry [H] and (d) traditional shops and services [S] and MODERN, containing (a) own business [OB], (b) self-employed including professionals[SP] and (c) workers including salaried persons [WS]. Under all these 7 types of vocations, 4-digit codes were generated to capture the precise nature of job at the best possible level of disaggregation (Dey and Banerjee 2016a, b).

With this exhaustive list of vocations generated from the second survey, a third survey has been conducted in January–February 2014 on the 240 households selected from these same 19 mouzas experiencing LOW or NO change in the pattern of land-use, to especially explore the nature of pull factors. An additional control was given to the average size of water bodies, which has an observed inverse relation with the possibility of change. Only those households were considered where at least some members are engaged in traditional activities. These 240 households supplied information on 980 individuals out of which 434 were engaged in some kind of economic activity according to the classification scheme prepared on the basis of the second survey. Information was collected on the socio-demographic background of the household members like age, education, their major and minor occupations, status of employment, adequacy of earning (at both individual level and the household level), willingness to switch vocation and the level of environmental awareness regarding the uniqueness of this eco-system. The average size of the

[5] 19 mouzas in all;

household turned out to be 4, 76% of working population is male, 30% of working members are within 30 years of age (younger), and 52% is either illiterate or having education up to primary level (low education). Coming to the type of vocation, 69.25% are solely engaged in traditional activities, 14.75% involved in purely modern vocations, and out of the remaining 16% engaged in mixed vocations the dominance of traditional vocation is noted in roughly two-third of the cases.

2.3 Vocation Switching Tendency

From the analysis of survey data, it is found that young age and formal schooling enhances the likelihood of vocation switching (Dey and Banerjee 2017). So, an in-depth investigation needs to be carried out into the likely process of shifting. If people find income inadequate for living, they are likely to diversify into other jobs and the switching process would be gradual. From water-based vocation like fisheries, they may avail initially other land-based options like agriculture, animal husbandry and finally move to services, both traditional and modern, of course, subject to their preparedness at the individual level. Four different types of traditional vocations have already been specified as F, A, H and S. In all 85.25% of the working population is engaged in traditional vocation with the relative share of A being 45.16% and that of F is 26.97%, H is 11.98% and S is 17.74%. Out of 26.97% engaged in F, 20.74% is engaged in pure F and the rest 6.23% is mixing vocation with other traditional activities. Similarly, for A the breakup is 31.34 and 14.82%. The duck rearing, poultry and animal husbandry (H) is generally taken up as a subsidiary engagement in conjunction with other major activities, both traditional and modern. Those who are in S are rarely mixing with other traditional vocations with almost a singular exception for A (only 4.38%). In fact, when people are diversifying to newer vocations and pursuing combination packages, they prefer modern alternatives. This result lends supports to the surmise of vocational transition from wetland or water-based livelihood practices to land-based activities, which would eventually break the correspondence between sewage water treatment and wastewater aquaculture. The presence of 64 individuals in purely modern vocation (14.75% of total working population) was not expected as the sample observations were purposively collected from households with engagement in traditional activities.

If this propensity continues and especially, if fisheries stop dominating the livelihood pattern, then that will not only affect the low-cost supply chain available to the city dwellers, will challenge the continuity of the waste management practice as well. Multiple agencies are active here with multiple stakes, which are at time working at cross-purposes leading to a classical case of coordination failure. An elaboration on this precise nature of agency conflict would be interesting to discuss.

2.4 Legal Initiatives to Protect EKW

The position papers of Calcutta Metropolitan Planning Organization (CMPO)[6] and that of the West Bengal State Planning Board (WBSPB) are evidence of strong objection to the eastward expansion of the city since early 1960s (Banerjee 2012). Filling up of water body was prohibited under the Town & Country Planning Act in 1979. In spite of that, the Salt Lake City was extended and the Eastern Metropolitan Bypass was constructed on reclaimed wetlands during the 1980s, making the core wetland area more accessible as well as vulnerable (Dembowski 1999). The Institute of Wetland Management and Ecological Design (IWMED)[7] was set up in 1986 with the primary objective of carrying out studies related to wetland; however, it was never given the statutory powers needed to play the role expected from it.

Thus, during the 1970 and 1980s there was a drastic shift in position at the level of urban development authority whereby the eastern fringe of the city suffered from a pre-dominance of unplanned, uncoordinated urban growth. The first major resistance came from the civil society in the year 1992. A pressure group called PUBLIC (People United for Better Living in Calcutta) filed a writ petition in the High Court to protect the EKW from urban encroachment. On September 24, 1992, High Court Justice Umesh Chandra Banerjee delivered the first major judgment on the matter in favor of protection of the core area of EKW. He ruled that *the wetlands were a gift of nature* and it is the court's job *to strike a balance between development and environment*. Court also mentioned that the wetlands were *too precious to be sacrificed for a mere township*. The exclusive protection of the core wetland area has one major drawback. Those eastern fringes of the city not covered by the High Court's order have since been exposed to rapid, inadequately planned urbanization. The buffer area to the wetland seems to have no control over encroachment, and this particular problem was first highlighted by a public interest litigation (Surojit Srimani vs. the State of West Bengal), which was filed in May 1995 to control Calcutta's eastern sprawl.

A landmark in the history of EKW conservation is its recognition as an international Ramsar site on the 19th of August 2002 (http://www.ramsar.org/pdf/sitelist.pdf). According to the Ramsar Information Sheet, '*the EKW is one of the rare examples of environmental protection and development management where a complex ecological process has been adopted by the local farmers for mastering the resource recovery activities.*' Following this international recognition, in the year 2006 East Kolkata Wetlands Conservation and Management Bill was passed and 12,571 hectares of land was brought within the wetland boundary. According to this bill, not only any new construction within EKW will be severely penalized but all existing constructions within this area would have to be demolished with immediate effect. In spite of these legal barriers, the attempt to encroach has become a perennial problem for EKW and the East Kolkata Wetlands Management Authority (EKWMA) was formed under the provision of the East Kolkata Wetlands (Con-

[6]Later merged with Kolkata Metropolitan Development Authority (KMDA);

[7]Later it was renamed and extended as Institute of Environmental Studies and Wetland Management (IESWM) in 2005;

servation and Management) Act, 2006, to resist these attempts. In 2011, the state government has developed a management plan and in 2012 a high-powered committee has been appointed to look into the matter.

However, the legal authority failed to understand the crucial role played by the buffer area (area between the main city and the wetlands) for the protection of the core area. For a metropolitan city which is growing leaps and bounds in the eastern side, the fringe area enjoys an enormous socioeconomic significance. If there exists no legal binding on the land-use pattern in buffer areas, it is practically impossible to resist any attempt of encroachment in the core area. The core area will experience different covert attempts of land conversion, and at a micro-level majority of small pond owners/fish growers will find the offer of land transfer financially lucrative enough to be acceptable. The problem lying in the legal sanction of this practice may be avoided by adopting different illegal practices. The procedure will follow a surreptitious path, but the ultimate outcome would be observable and overt. Whether a mere legal protection to wetland alone is adequate for preserving the eco-system-based livelihood in EKW is our primary concern. What is most important in this context is the perception of the local residents. If the local people are confident about the strength of this legal protection, then they would like to continue in their traditional vocation and the livelihood would be protected. However, when the legal standing is only a formal façade which is not backed by public confidence, everyone would be interested in switching to the modern vocation to reap the advantage of the newly created opportunities. Here, legal provisions alone would fail to protect the eco-system.

2.5 A Story of Coordination Failure

The regulator is enacting different legal provisions for the protection of land-use pattern, which are creating confusion among the local people and in the absence of strong political will nothing is getting enforced with appropriate thrust. There is not much coordination between KMC and KMDA, where the former relies on EKW for cost-free natural treatment of wastewater and the latter is in charge of urban expansion and development. In fact, multiple stakeholders starting from KMC, KMDA, HIDCO, NGOs, Fishery cooperatives, local residents, eco-tourism resort owners, real estate developers, and finally wetland institutes like IWMED, EKWMA, with their multiple agenda regarding the wetlands are not working in harmony. An absolute coordination failure occurred due to this complicated web of associations among stakeholders documented in Table 3.

The lack of understanding and transparency is hindering the system to achieve its intended sustainability. Everyone is contemplating corrective moves from his/her own perspective, and in the process, the retarding influences on others nodes are passing unnoticed. Though the East Kolkata Wetland Management Authority has evolved to manage this complex and challenging issue, the institution does not enjoy any statutory power till date to ensure effective coordination. The effective interven-

Table 3 Identification of stakeholders and the coordination problem

Body	Perspective	Dept./party	Purpose/ responsibility	Challenge	Instrument	Problems
Govt.	Public management	Irrigation & waterways	Maintaining canals which carry the sewage water into the wetlands	Maintenance	Routine Govt. budget	Dredging done for main canals only, smooth flow of wastewater *(input in fishery)* into branch canals hampered
		KMDA	Decentralized planning & development across the urban and rural areas	Development	Planned expenditure	Planning did not take the sewage management issue that may arise from urban invasion into the wetlands
		HIDCO	Plans and executes development projects in the entire Rajarhat Area, Kolkata, West Bengal			Urban expansion at an immediate vicinity i. changed the landscape ii. Stakeholders of EKW Ramsar conservation site in no way received benefits of New Kolkata Township, either by selling out their property or in terms of getting improved urban amenities

(continued)

Table 3 (continued)

Body	Perspective	Dept./party	Purpose/ responsibility	Challenge	Instrument	Problems
		Panchayat	Pivotal agency for unleashing comprehensive rural development	Local governance	West Bengal Panchayat Act, 1973	Non-specific
		KMC	Manage wastewater of Kolkata		West Bengal Municipal Act, 1993	KMC has no administrative power i. to maintain canals and sub canals ii. to protect fishery-based livelihood, crucial wastewater treatment No sufficient financial provision for sewage treatment available so far
	Wetland management	EKWMA	Maintain the existing land-use practices along with its unique wise use practices (According to Ramsar Guidelines)	Conservation and protection	East Kolkata wetlands (Conservation and Management) Act, 2006	Lack of statutory power to stop land conversion
		IWMED	Carrying out studies related to wetland functions & its ecology, wetland mapping etc.	Awareness building	Documentation & research	Carry out research only, no role in protection

(continued)

Table 3 (continued)

Body	Perspective	Dept./party	Purpose/ responsibility	Challenge	Instrument	Problems
Non-Govt.	Civil society	NGO	Mobilization of local awareness at grass root level	Conservation & resistance	Awareness camps and formation of local groups	Have some opinion regarding the conservation issue; however, the legal authorities can take action in 'curative' way not 'preventive'
		Local residents		Conservation and protection		
Economic group	Income generation	Local residents	Livelihood protection	Protection of private property right	Enhanced transparency in documentation	Newer generations not inclined to traditional livelihood are keen on sale of protected land though illegal means
		Fishery cooperatives	Siltation & lease renewal	Stop land filling & ensuring profitability of fishermen	Cooperative resistance	Irregularities in registration procedure
		Real estate developer	Land conversion & acquisition	Land filling, changing signature of landscape & profit earning via sale of protected land	Illegal means	Land transfer & sale of land which is apparently protected area
		Eco-tourism	Commercial interest by protecting the landscape	Prevent land filling & siltation in respective eco-tourism hubs	Promotion of waterfront activities	Only the landscape can be protected; however, protecting '*wise use*' in a holistic manner is not possible

Source Dey and Banerjee (2015)

tion and related design of implementation mechanism on the part of the designated high-powered committee on the conservation of the EKW is yet to claim any strategic move. Of course, time is still young to take any conclusive position in this regard.

The impression that we have gathered from the foregoing discussion is that EKW is about to lose its ecological integrity, which mainly refers to the long-term health of the system in terms of interactions among the physical, chemical and biological elements of the eco-system creating favorable social conditions for sustaining ecological basis of human life. Here, none of the stakeholders are approaching the problem from an integrated, comprehensive perspective.

The civil society is expressing concern over protection of wetland, where that step alone would be insufficient for the conservation of the eco-system.

3 Study III: Solid Waste Management[8]

Recognition of the social costs associated with traditional practices of urban waste management in India led to the formulation of Municipal Solid Waste Management and Handling Rules (2000). However, compliance with the proposed collection and disposal involves higher commitment in terms of both time and money on the part of the residents, local bodies as well as the state and the central government. This is an interactive framework where each stakeholder has some definite role to play. If the beneficiaries decline to play the role expected from them, then the entire approach will fall flat. So, some information about the city dwellers' perception regarding the value of the environmental improvements conferred upon them through the implementation of the new rule would be important for the planners to know. If the residents do not consider the project worthy and decline to separate waste at source, the subsequent processing plan will not work. Given the non-market characteristic of waste disposal services, we infer about beneficiaries' perceived demand for the proposed service by means of Double-Bounded Dichotomous Choice Contingent Valuation Survey conducted in 2006 in the Bally Municipality of the district of Howrah, West Bengal, before the adoption of the waste management program proposed by the MSWMHR. We estimate the average WTP by controlling for anchoring bias and use the annualized value of cost to examine the feasibility of the proposed system. The system turned out to be feasible. It was introduced in 2008, and part of our survey area was covered by it.

An attempt has been made next to identify the factors that influence perception of program benefit of the recipients when a hypothetical public program is implemented, at least partially. When the hypothetical program becomes real, actual experience of program features is likely to update the information set of the recipients. In fact, the recipients would now have a clear assessment of the benefits associated with each component of the waste management program like garbage collection and disposal.

[8]The discussion presented in this section draws heavily from Sarkhel and Banerjee (2010) and Sarkhel et al. (2015); necessary permissions for re-use have been obtained.

The perceived net benefit that influenced his WTP in the pre-program stage would accordingly get revised. Furthermore, the recipient's level of satisfaction with the service delivery is likely to play an important part in this respect. We compare pre- and post-program willingness to pay (WTP) estimates for improved waste management in Bally Municipality, India, by conducting another survey on the same set of households in 2011. Surprisingly, it was found that the post-program predicted WTP falls by more than 50% even when there is a common perception about substantial improvements in the quality of urban environment. Reasons behind this apparently puzzling situation were studied from two angles: In the demand side, there is a possibility of revision in the actual cost of sacrificed leisure needed to participate in the improved waste management program. Another explanation may come in terms of a mismatch between expected and offered service attributes causing disutility to dampen households' perceived value of the program benefits. In the supply side also this reduction in WTP signals some important messages: It might act as an indication to the local bodies regarding the required quality of service and the scale of outreach as the expansion of the program needed to finance the operation and maintenance expenses by supplementing the property tax bases through user fees. If outreach expands the average unit cost would fall and with lower WTP, the program will break even.

3.1 Bally Municipality

The Bally Municipality with a population over 0.261 million is located in the Gangetic plain of West Bengal in the district of Howrah in India. Traditionally, Bally developed as an industrial area along the river plain of Ganges. As a result, it has a large number of migrant populations from other states of India. These people mostly worked in informal industries as well as formal ones like jute and metal industries. In fact, there has been a significant growth of population in Bally during 1991–2001 amounting to more than 40%, and this has intensified the congestion in the municipal area spread over 11.81 km^2. The provisioning of waste management services in Bally Municipality displays all the traits that are typical of an urban agglomerate of developing countries. According to official estimate, 150 tons of waste is generated per day in the Bally area most of which comes from the residential households. The collection activity is irregular and the uncollected waste putrefies in roadside vats imposing severe health and aesthetic cost on the inhabitants. According to municipal officials, door-to-door collection of household waste is partially implemented covering a little more than 30% of the total population.

On the disposal side, the local body is fast running out of its existing dumping ground and given the fact that Bally has the highest decadal urbanization rate in the district, the cost of setting up new grounds is bound to escalate in future. The scarcity of land necessitates the adoption of alternative disposal practices along with landfill that would divert maximum waste from the dumping ground and at the same time would be less land intensive. Here, composting is considered to be one such feasible

option. However, one major impediment toward its successful adoption is the fact that people separate only those parts of the waste stream that they can trade with the itinerant buyers. This removes components like paper, glass and even batteries from the daily waste but may leave out harmful contrary materials like paints, expired medicine and soggy plastics that might prohibitively increase the sorting cost of organic waste needed for the production of standardized quality of compost. However, from the point of view of the municipality, composting seems to be the most relevant option for other reason also: There is a potential for converting waste into wealth and consequently a scope for cost recovery as the sale of organic manure can also bring in some revenue. With all the preconditions in favor of launching the project, the long-run sustainability of the proposed garbage disposal scheme crucially hinges on the perceived benefits of the households from the same.

3.2 Contingent Valuation Survey

With improvement in the quality of living environment, our sense of well-being enhances. It is quite common to observe individuals to express strong preferences for clean water, fresh air, wider road, quiet neighborhood and so on. Some of these amenity services are marketed and some are not. The goods for which there is no well-defined market, one way to understand the preference of the potential beneficiaries is to make use of survey instruments and ask them directly to state their preference regarding the provision of the good. Contingent Valuation Survey is one such stated preference method where in the absence of any possibility of market signal to reveal the stakeholder's preference for the service in question, direct bids are elicited to assess people's willingness to pay(WTP) for it. In our context, municipal solid waste management is an urban amenity in the sense that with proper management practices significant benefits in health and aesthetics can accrue to the city dwellers. As a result, welfare of society at large increases. Hence, CVS method can be applied hereby projecting the possible improvement in the living condition of the respondents and asking her about the extent of contribution she is willing to make to enjoy the benefits of such change.

A particular feature of Bally Municipality is the abundance of slum population in the area, that is, 27% of the total population, and their inclusion in the hypothetical market is vital for offering them the opportunity of a better quality of life from the public policy point of view. Keeping this in mind, we allocated the 29 wards in three different strata with wards having high, medium and low slum population. From each stratum, 2 wards are selected with purposive randomness. Bally Municipality has three broad administrative zones, viz. Bally, Belur and Liluah. Hence, we picked up wards from all three zones and three categories. From each ward, we randomly selected households on the basis of household list provided by the municipality that came to a total of 570 survey respondents. Out of these 570 households, only 496 could be identified five years later, at the time of the repeat survey.

Typically, a Contingent Valuation Survey concocts a hypothetical market for the good to be valued and elicits the amount that an individual is willing to trade in exchange of that. In the process, there are three important design issues that are to be dealt with (1) accuracy in the program description that is to be offered, (2) constructing market institutions that involves informing the respondent clearly about the service provider and the payment vehicle and (3) designing value elicitation questions for estimating average WTP of the sample respondents. So far as the program description is concerned, we emphasized on two aspects: replacement of open dumps by sanitary landfills and initiation of compost production by utilizing the biodegradable portion of daily household waste. However, it was also mentioned that for the latter daily collection of source-separated waste would be introduced whereby households would be required to segregate the dry and wet waste in containers provided by the municipality. Respondents were informed that the subscription for the offered program would be collected as a monthly tax along with other utility bills by the Municipal authority. We designed a double-bounded dichotomous choice framework that posed an initial and follow-up question to the respondents. Three starting bids of Rs. 5, 10 and 20 were randomly allotted to the sampled households: 'Yes' response was followed by a doubling and 'No' response by a halving of the asked bid.

3.3 Findings of Interest

Out of 496 surveyed households, only 62% received the new package where the households are provided red and green disposal bins, and there was doorstep collection of source-separated wastes. The remaining 38% continued with an arrangement where there are no source separation requirements. Thus, there is a scope to observe variation in both WTP and attitude of the households over time as well as across groups. However, descriptive analysis across program recipient and non-recipient reveals that even the program non-recipients engage in waste segregation and recycling and there is no statistically significant difference between them in terms of time required for the segregation or number of items recycled. This indicates a strong presence of information externality within the locality.

After program implementation, both the recipients and the non-recipients report enhanced usage of municipal disposal service. There has been an overall increase in tendency to dispose garbage in municipal vans (an increase from 44.85% in 2006 to 81.45% in 2011). In keeping with this, almost 88% of the sampled households ($n = 436$) in 2011 agreed to avail the service against a monthly payment. In fact, 84% of households who denied accepting the paid service in 2006 have now agreed to subscribe even when they are not the direct program recipient. Even across recipients and non-recipients, the WTP for the current service is more or less the same. Against this huge support for program continuance and adoption, the stated WTP has gone down, on an average, in the post-program period (from INR 17 per household per month in 2006 to INR 12 per household per month in 2011), and here also no

significant difference is noted in the average bid amount for program recipient and non-recipient. In fact, segregation time per unit of garbage has fallen in the post-program period. This confirms our hypothesis that assessment of time requirement was prone to be an over estimation in the hypothetical stage. In the later round, the program has actually been launched and the respondents are more certain about the actual time commitment for in-house source separation and recycling of wastes. The reassessment of time requirements is also associated with significant changes in recycling levels in the post-program period. A representative household now recycle almost twofold number of items than he used to do in the pre-program period, and there is an increase in both extent and intensity of recycling. Thus, in the post-program phase there are concrete evidences that benefit accrue via two channels: First, there had been improvements in the overall cleanliness in the area, and secondly, recycling activities have increased substantially, which contributes toward reduction in net increase in the final waste generated per unit of household consumption even when the actual level of consumption (in terms of monthly expenditure) has gone up significantly.

Finally, we look at households' satisfaction with the components of waste management services that have been implemented. We take into consideration the households' choice about four specific program components like storage of garbage, time of collection, preference of service provider and method of disposal. Satisfaction indicators have been constructed as: (i) satisfied with status quo, (ii) dissatisfied with any one dimension, (iii) dissatisfied any two dimensions and (iv) dissatisfied with more than two aspects. The households' distribution over these four categories indicates that most of the households differ from status quo in respect of one dimension, and only 28% of the sampled households prefer status quo with respect to all attributes. Estimates of predicted WTP reveal that average WTP in 2011 (INR 12) is substantially lower than average WTP in 2006 (INR 34). The post-program predicted WTP is even less than half of its value in the pre-program phase. Apparently, the lower values of WTP come as a surprise as net benefit of the households has most likely increased due to overall improvement in the municipal environment as well as private benefits accruing due to enhanced cleanliness of his immediate neighborhood.

4 Overall Assessment

Three studies related to micro-level intervention toward participatory development management have been presented here, and the only common point among them is some unexpected outcome produced by the process. The contexts are widely different though the problem faced in each case is indicating a bottleneck in ensuring some attainable non-empty core. Our experience is not singularly different from that of other researchers in the field. By analyzing the cases from West Africa and India, it has been shown by Beck and Nesmith (2001) that instead of economic success in managing common property resources to provide support to the local people focus should be laid on sustainable livelihood. This approach would protect their identity and cul-

ture, and would help to organize the development projects on more equitable basis, both socially and economically. Kumar (2002) has shown strong 'have'-preference of India's JFM over the last 40 years leading to further marginalization of the 'have-nots.' In fact, the most serious critique of participatory approach to development in terms of some built-in conceptual paradox has been advanced by Mohan and Stokke (2000) and Cleaver (1999). The former one indicates the inductive localization of development interventions in case of participatory management, which is overshadowing the holistic approach in a broad macro-theoretic perspective. The latter one highlights the possible snags faced by these project centric micro-development initiatives. Participation is claimed to enhance efficiency of outcome (economic gain), leading to greater equity and empowerment of the vulnerable group (social gain) and through the conservation of local resources, traditional knowledge and culture makes the development outcome more resilient to ecological shocks (environmental sustainability). For economic and environmental gains, participatory management is a *means*, whereas for the social gain it is an *end* in itself. If these three aspects work in synergy, there is no problem. However, such instances are rare and exceptional, not rules.

The success of the participatory approach crucially depends on the evolution of appropriate social institutions that would help to formalize mutual expectations of cooperative behavior and reduce the cost of individual transactions by working as an insurance against the problem of opportunism and free riding. So, the effective institutions are necessarily formal, whereas the traditional institutions are essentially informal in nature. They are not governed by contracts, associations, committees and clearly defined property rights which are generally considered as effective instruments for reducing transaction costs.

References

Banerjee, S. (2012). Overseeing the disappearance of an ecosystem: Governance in West Bengal and Wetlands to the East of Kolkata. In J. Bandyopadhyay, K. Chopra & N. Ghosh (Eds.), *Environmental governance: approaches, imperatives and methods*, INSEE (pp. 190–222). New Delhi: Bloomsbury.

Banerjee, S., & Dey D. (2017). Eco-system complementarities and urban encroachment: A SWOT analysis of the East Kolkata Wetlands, India. In *Cities and The Environment, Supported by Loyola Marymount University, Bepress* (Vol. 10(1)). USA: Berkeley Electronic Press, July 2017 (ISSN: 1932-7048).

Beck, T., & Nesmith, C. (2001). Building on poor people's capacities: The case of common property resources in India and West Africa. *World Development, 29*(1), 119–133.

Berkes, F. (2003). Alternatives to conventional management: Lessons from small-scale fisheries. *Environments, 31*(1), 5–19.

Bit, J., & Banerjee, S. (2015). Joint forest management in Birbhum forest division, West Bengal: An assessment with multi-criteria decision analysis. In *The survey, IISWBM* (Vol. 55, Nos. 3&4).

Brett, E. A. (2003). Participation and accountability in development management. *The Journal of Development Studies, 40*(2), 1–29.

Castro, P. C., & Nielsen, E. (2001). Indigenous people and co-management: Implications for conflict management. *Environmental Science & Policy, 4,* 229–239.

Chopra, K., & Kadekodi, G. K. (1991). Participatory institutions: The context of common and private property resources. *Environmental & Resource Economics, 1,* 353–372.
Cleaver, F. (1999). Paradoxes of participation: Questioning participatory approaches to development. *Journal of International Development, 11,* 597–612.
Dembowski, H. (1999). Courts, civil society and public sphere environmental Litigation in Calcutta. *Economic and Political Weekly*, *34*(2), 49–56.
Dey, D., & Banerjee, S. (2013a). Change in land-use pattern of East Kolkata Wetlands: Concern and consequences. In *Urban India, National Institute of Urban Affairs (NIUA)*, January–June 2013, (Vol. 33(1), pp. 154–174). (ISSN: 0970-9045).
Dey, D., & Banerjee, S. (2013b). Ecosystem & Livelihood Support: The Story of East Kolkata Wetlands. In *Environment and urbanization Asia SAGE on behalf of National Institute of Urban Affairs (NIUA)*, October 2013, (Vol. 4 (2), pp. 325–337). (ISSN: 0975-4253).
Dey, D., & Banerjee, S. (2015). Role of legal institutions in managing urban ecosystem: A case study of East Kolkata Wetlands. *CBS Journal of Management Practices*, *2*(1), 41–58. (ISSN 2349-5073).
Dey, D., & Banerjee, S. (2016a). Disposal of wastewater and the east-bound growth of the City of Kolkata: A compatibility analysis. In *Environment, development and sustainability*. Springer, July 11th, 2016. (ISSN 1387-585X, https://doi.org/10.1007/s10668-016-9835-2).
Dey, D., & Banerjee, S. (2016b). Land-use change and vocational transition in east Kolkata wetlands: evidence from time-diary. In *Environment and urbanization Asia*, *SAGE on behalf of National Institute of Urban Affairs (NIUA)* (Vol. 7(2), pp. 243–266). (ISSN: 0975-4253).
Dey, D., & Banerjee, S. (2017). How expensive is the decay of East Kolkata Wetlands? An estimation of opportunity cost for Kolkata. In J. Mukherjee (Ed.). *Sustainable urbanization in India: Challenges and opportunities*. Springer. ISBN978-981-10-4932-3.
Diaz-Balterio, L., & Romero, C. (2008). Making forestry decisions with multiple criteria: A review and an assessment. *Forest Ecology and Management, 255,* 3222–3241.
Forman, E., & Peniwati, K. (1998). Aggregating individual judgments and priorities with the Analytical Hierarchy Process. *European Journal of Operational Research*, *108*, 165–169.
Ghosh, D. (2005). In *Ecology and traditional wetland practice: lessons from wastewater utilization in the East Calcutta Wetlands*, Worldview, India.
Gilbert, A. (2006). Good urban governance: evidence from a model city? *Bulletin of Latin American Research, 25*(3), 392–419.
Kangas, J., & Kangas, A. (2005). Multiple criteria decision support in forest management—the approach, methods applied, and experiences gained. *Forest Ecology and Management, 207,* 133–143.
Khadka, C., &Vacik, H. (2012). Use of multi-criteria analysis (MCA) for supporting community forest management. In *iFores—Biogeosciences and Forestry*, *5*, 60–71.
Kumar, S. (2002). Does "Participation" in common pool resource management help the poor? *A Social Cost-Benefit Analysis of Joint Forest Management in Jharkhand, India., 30*(5), 763–782.
Mendoza, G. A., & Martins, H. (2006). Multi-criteria decision analysis in natural resource management: A critical review of methods and new modelling paradigms. *Forest Ecology and Management, 230,* 1–22.
Mendoza, G. A., & Prabhu, R. (2005). Combining participatory modelling and multi-criteria analysis for community-based forest management. *Forest Ecology and Management, 207,* 145–156.
Mohan, G., & Stokke, K. (2000). Participatory development and empowerment: The dangers of localism. *Third World Quarterly, 21*(2), 247–268.
Paul, S. (1989). Poverty alleviation and participation: The case of government-grassroots agency collaboration. *Economic and Political Weekly, 24*(2), 100–106.
Saaty, T. L. (1980). *The analytic hierarchy process*. New York: McGraw-Hill.
Saaty, T. L. (2008). Decision making with the analytical hierarchy process. *International Journal of Services Sciences*, *1*(1), 83–98.
Sarkar, R., & Sinha, A. (2015). *Another development*, Routledge.

Sarkhel, P., & Banerjee, S. (2010). Municipal solid waste management, source-separated waste and stakeholder's attitude: A contingent valuation study. *Environment, Development and Sustainability, 12*(5), 611–630.

Sarkhel, P., Banerjee, S., & Banerjee, S. (2015). Willingness to pay before and after program implementation: The case of municipal solid waste management in bally municipality, India. *Environment, Development and Sustainability*. https://doi.org/10.1007/s10668-015-9659-5, ISSN: 1387-585X.

Yan, J., Wang, R., & Wang, M. (1998). The fundamental principles and ecotechniques of wastewater aquaculture. *Ecological Engineering*, *10,* 191–208.

Sustainability as Corporate Strategy: Importance of the Values of Ecosystem Services for Businesses

Nilanjan Ghosh

1 Introduction

Sustainability, as a concept and in practice, is increasingly being acknowledged by businesses across the world as part of their core operational strategy, and is being prominently adopted for market positioning and branding. Though apparently this trend is more prominent in the developed world, many businesses in the developing world are also catching up (Ghosh et al. 2016). Yet, a large component of business enterprises especially in the developing and the less developed world still remains ignorant of the significance of "sustainability" as a corporate strategy. In the process of protecting their own bottom-lines, the questions that loom large before them are: Why should businesses bother about sustainability? Does it affect their profitability?

Despite the growing recognition of sustainability in some domains of business, the concept remains abstract and theoretical for many businesses. At a macro-policy level, the recognition of sustainability as a triad or trinity of economic, social, and environmental concerns that need to be reconciled (see Fig. 1) is more ubiquitous in paper, though less in practice. At the firm level, the Brundtland Commission definition of sustainable development as "development that meets the needs of the present without compromising the ability of future generations to meet their own needs", is often treated as an "ethical", "normative", and a "value judgmental" concern. That is precisely why even today, sustainability is treated by a large component of the business as a "social responsibility".

However, at a micro-level, organisations accept that protecting its capital base is important for its own bottom-line. Yet, they hardly recognize the possibility of

N. Ghosh (✉)
World Wide Fund for Nature, New Delhi, India
e-mail: nilanjan.ghosh@gmail.com

N. Ghosh
Department of Economics, Observer Research Foundation, Kolkata, India

P. Ray et al. (eds.), *Economics, Management and Sustainability*,
https://doi.org/10.1007/978-981-13-1894-8_10

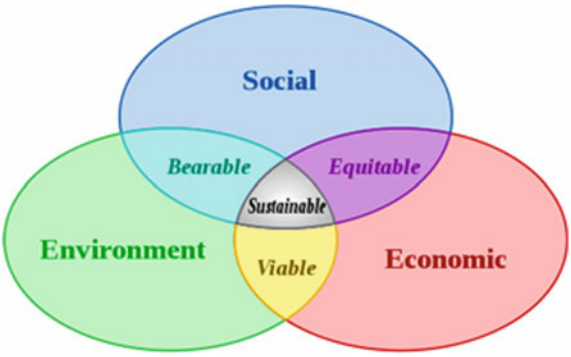


Fig. 1 *Source* https://en.wikipedia.org/wiki/Sustainable_development

extending the notion to the world's natural and human resources. In fact, till a point in time, environmental and ecological concerns were perceived in most parts of the world as an ethical argument which might be anti-developmental (Krehmeyer et al. 2010; Ghosh 2015). This entire feeling made conservation initiatives move asymptotically with the core human business, and was initially taken as something external to the fundamental human existence.

In this paper, I talk of sustainability primarily from the perspective of conservation. It is in this context, I bring in the notion of Creating Share Value (CSV), as conceived by Porter and Kramer (2006 and 2011) in a broader context, as also the importance of ecosystem services and their valuation in shared value creation. As such, in course of arguments, it will emerge that by embracing sustainability as corporate strategy, firms are essentially creating shared value. This is exhibited in this paper through the application of valuation of ecosystem services. In the process, the ensuing discussion also entails how the firms need to use these values of ecosystem services in the course of decision making.

In Sect. 2, I talk of the changing paradigm from looking at conservation needs as moral-ethical concerns to looking at conservation as selfish human needs. The role of ecosystem services (i.e. services provided by ecosystem to human society for free) is crucial here. Section 3 talks of the CSV notion, and how it is embedded in sustainability as corporate strategy. Section 4 talks of the importance of valuation of ecosystem services, and exhibits two cases of ecosystem service values at two different scales: one at the scale of a wetland ecosystem, and the other at the scale of a landscape. In this context, it also brings in the notion of ecosystem services as "GDP of the poor" thereby linking ecosystem services to livelihoods. Section 5 consists of the concluding remarks highlighting the significance of conservation initiatives by the firms for their own long-run sustainability. In the process, it discusses how firms may use the ecosystem service values for the cost-benefit analysis of projects that entail their interventions in the ecosystem. Thus, it argues that while embracing sustainability as corporate strategy is about creating sustainable bottom-lines with the firm's embedment in the broader ecosystem, valuing ecosystem services and integrating them in the decision-making is about creating shared value.

2 The Changing Perception

A few decades ago, "conservation as an anti-developmental force" was also the feeling in the developed world. But, the accrual of knowledge and scientific understanding at the interface of nature, economy, and society ever since the 70s started changing this perception in the developed world. The extensive development of science in this domain also made humans understand that there is bidirectional causality between ecosystem and the economy. Ever since the Club of Rome's prediction of apocalypse in their *The Limits to Growth* thesis (Meadows et al. 1972), the human response to the "approaching doomsday" has been characterized by extensive research, gradual knowledge accrual through global assessments, and conventions.

The Earth Summit of 1992 adopted the Brundtland Commission Report's definition of "sustainable development", and opened the Convention of Biological Diversity (CBD) for signature. The CBD became effective from December 1993. With CBD, for the first time, the framework of international law recognised conservation of biological diversity as an integral part of the development process.

On the other hand, when David Pearce and Kerry Turner talked of their magnum opus *Circular Economy* in 1989, the term caught up like wildfire in the broader discourse of the interactive dynamics of environment and development. Circular economy marks a clear departure from the very reductionist linear growth thinking of "take, make, dispose" to a more holistic paradigm that conceives of the economy as being embedded in the ecosystem (Pearce and Turner 1989). Therefore, the bidirectional causalities between the two forces (economy and ecology) are better acknowledged.

The *Millennium Ecosystem Assessment* (MA) of 2005 enhanced human understanding of the fact that the ecosystem functions in its own inimitable ways to provide ecosystem services (benefits) to the human society in the form of provisioning services (e.g., food, raw materials, genetic resources, water, minerals, medicinal resources, energy, etc), regulating services (e.g., carbon sequestration, climate regulation, pest and disease control, etc), cultural services (tourism, religion, etc), and above all, supporting services that are necessary for production of all other ecosystem services (e.g. nutrient recycling, gene-pool protection, primary production, soil formation, etc).

The economy-ecosystem linkage seemed much clearer with the better delineation of ecosystem services. Referring to Fig. 2, the social system creates economic forces and enforces economic actions due to the mismatch between needs and availability. "Actions" on the ecosystem are fundamentally meant to satisfy economic needs: Hence "Actions" are "Economic Activities". "Economic Activities" emerge as interventions on the ecosystem structure and functions, and provide ecosystem services to human society. The proper delineation of "ecosystem services" helped in understanding the direct linkage between human society and biodiversity: for every bit of existence of human society, there is a critical need for the biodiversity as a "stock" to exist, to ensure the "flow" of these ecosystem services! Recent scientific assessments at this interactive interface like *The Economics of Ecosystems and Biodiversity*

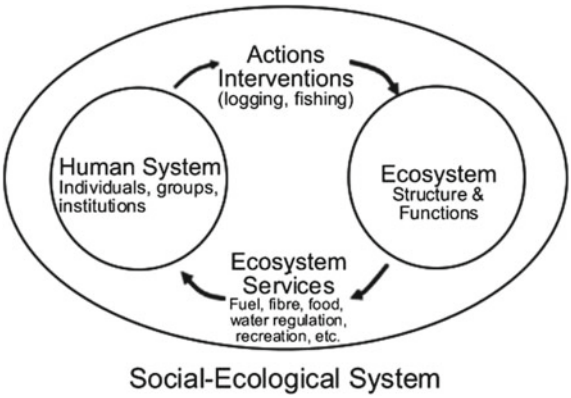


Fig. 2 The social-ecological system

(TEEB), published in 2010, recognised that these ecosystem services are the "GDP of the poor", as the poor's incomes and survival are dependent on the ecosystem.

While recognising the importance of the food-chain in the context of the ecological balance so as to ensure the integrity of the ecosystem structure and functions in order to ensure the flow of ecosystem services, conservation goals become important. For businesses to survive, natural resources are needed. For sustainable management of the natural resources like forests, wetlands, rivers, etc. one needs to set the right conservation goals for flora and fauna, which through their natural functioning, support and sustain these resources, and provide ecosystem services.

Yet, there is no harm reiterating that till a point in time the businesses felt that the conservation needs of the planets do not really concern them. The entire environmental concern was not intrinsic to its core business and business plan. Pollution is an externality, and often government regulations force them to internalise costs of pollution. There was hardly a recognition that businesses inextricably depend on biodiversity through a well-defined supply-chain, whose recognition is obscure in the public domain, especially in India and the developing world. Later on, however, this dependence has been adequately recognised by the World Council for Sustainable Development in various publications (e.g. WBCSD 2008, 2013).

2.1 The WBCSD and Natural Capital Protocol

The very establishment of World Business Council for Sustainable Development (WBCSD) in 1995 bore ample testimony of the fact that businesses need to acknowledge their dependence on ecosystems for their functioning and existence, and the non-acknowledgement of this linkage can only lead to the extinction of the raw material base. The WBCSD is a CEO-led, global advocacy association of 200-plus international companies dealing exclusively with business and sustainable development.

Lately, the recognition of biodiversity conservation has become extremely important from the perspective of treating the biodiversity as "natural capital". Natural capital is the world's stocks of natural assets which include geology, soil, air, water and all living things. It is from this natural capital that humans derive the ecosystem services. The term "capital" therefore apparently has the same connotation as it has in traditional economic thinking where capital is thought to be an important factor of production: "natural capital" is the critical and necessary factor in producing ecosystem services. While new investment can lead to addition in capital stock thereby raising production, investment in "natural capital" can help in sustaining the good health of the ecosystem and its services.

At the same time, there is a need to produce comparable information on natural capital for better decision making so as to measure, value and manage natural capital as carefully as is done for financial capital. This will enable business to make a meaningful contribution to business, society and the environment. It is in this context that the Natural Capital Protocol has been initiated by the Natural Capital Coalition, which is a collaboration of the world's leading institutions from business, science and academia, membership organizations, standard setting, finance, policy and conservation, subscribing to a common vision of conservation and enhancement of natural capital through initiatives by business.

The Protocol entails a standardized framework for identification, measurement, and valuation of the ecosystem services, thereby helping the firm understand its dependencies on nature and also the impacts of its initiatives on the ambient environment and the biodiversity. In the process, the Protocol helps inform business decision-making and presents the first step towards a consensus on a universal approach. The Natural Capital Protocol Toolkit sorts through the wealth of tools, methodologies and approaches available for natural capital measurement and valuation and maps them against the Natural Capital Protocol framework.

3 Embracing Sustainability and Creating Shared Value

The myopic perception of the corporate sector about the ecological concerns began changing with the notion of "sustainable business", as they started understanding that their longer term survival is contingent upon the resource base that is steadily getting depleted; their long-term costs will only increase as they impede on the working of the ecosystem thereby diminishing the capacity of the ecosystem to provide its services; those with a "sustainable" and "green" business model will generally have a competitive advantage, and will earn their local community's goodwill and see their efforts reflected in the bottom line. Businesses, being an integral component of the broader social-ecological system, are inherently dependent on the "natural capital" for their short-run and long-run survival.

As such, many businesses view their principal objective as making money. Others recognize a broader social role. Somehow, the lacking consensus between business leaders on this ground in the developing world is ubiquitous. Businesses continue

to face the trade-off between the need for making money for their survival, while searching for projecting their social face. There are many who are completely oblivious of their social role, and constant reiterations in various forums fail to infuse that culture. The concern of crony capitalism is not yet offset by social initiatives by businesses in India and south Asia: business has been criticized as a major cause of social, environmental, and economic problems; companies are widely thought to be prospering at the expense of their communities; at times trust in business falls to such a nadir that governments' corrective policies undermine competitiveness and sap economic growth.

Therefore, when viewed through the sustainability lens, the trade-off for the firm also arises between the choices of short-term rent-seeking behaviour, and long-term optimization. Porter and Kramer (2006 and 2011) have argued that firms' focus on optimizing short-term financial performance leads them to overlook the greatest unmet needs in the market as well as broader influences on their long-term success. Why else would companies ignore the well-being of their customers, the depletion of natural resources vital to their businesses, the viability of suppliers, and the economic distress of the communities in which they produce and sell? It is in this context that Porter and Kramer bring in the notion of CSV or "creating shared value"—a process by which companies could bring business and society back together by generating economic value in a way that also produces value for society by addressing its challenges. A shared value approach reconnects company success with social progress.

According to Porter and Kramer, firms can create shared value in three distinct ways: by re-conceiving products and markets, redefining productivity in the value chain, and building supportive industry clusters at the company's locations. This can be exhibited in the case of Nestlé, who being players in the global value chain in their various products, attempted to redefine their procurement process in coffee. They worked intensively with small-holder farmers of the poverty-stricken areas of the developing and less-developed world, who were trapped in the vicious cycle of low yield, poor quality, and environmental degradation. Nestlé joined hands with the farmers by advising them better on farming practices, helped growers secure inputs at better cost, and started offering premium for better qualities. Therefore, quality certification and better market linkages with rationalization of the supply-chain helped yields and quality increased the growers' incomes, the environmental impact of farms shrank, and Nestlé's reliable supply of good coffee grew significantly. This entire initiative created shared value which was embedded in the firm's strategic subscription to principles of sustainability.

Lately, by embracing sustainability, and by addressing social, economic, and environmental concerns, some Indian firms are in the process of creating shared value. With the recognition that in the long-run, the corporate sector is dependent on ecosystem services provided by nature like all human endeavours, and that such services affect the bottom-lines of the organisation in the long-run, sustainability is being increasingly viewed as part of corporate strategy. The Aditya Birla group and Tata Power are important cases in this regard (Ghosh et al. 2016).

4 Understanding the Importance of Ecosystem Services

Given this background, it is important that the corporate sector understand the important role of ecosystem services for their own business and for the survival of the global community at a bigger scale. At the very next stage, it is important to understand the footprint created by their own initiatives:—this entails understanding the ecosystem service losses created by the projects or initiatives. This is where the importance of natural resource accounting enters the scene. Essentially, this entails putting a monetary value to ecosystem services (Bockstael et al. 2000; Costanza et al. 1997, 2014).

Valuation of ecosystem services becomes important in various respects. Firstly, a monetary value to ecosystem services helps the community understand their importance to human society (Chopra and Adhikari 2004). Secondly, valuation offers a somewhat objective instrument for decision making. Thirdly, valuation of ecosystem services can raise awareness of the market and the policy-makers on the importance of the ecosystem services under consideration. Fourthly, ecosystem service valuation can help legal proceedings determine damages where a party is held liable for causing harm to another party: Pollution from upstream areas affects the downstream ecosystems negatively. To deal with compensation policies properly, the economic value of the harm so caused needs to be assessed to obtain the extent of the negative externalities. Fifthly, valuation of ecosystem services can help revise investment decisions: e.g. infrastructure development, that might otherwise ignore the related harm expected to be caused to the natural environment and consequent loss to the ecosystem services.

The firm's intervention in such a scenario can happen in the following way:

1. The firm needs to see to what extent its working or new projects intervene into the ecosystem;
2. In the process, it needs to hire or give contract to a consulting economist to assess the value of the loss in ecosystem services that its intervention may cause;
3. For the purpose of sustainability, it needs to create processes to offset the negative impacts. Valuation of ecosystem services can again play a role for monitoring the change.

As such, companies like Tata Power and Hindustan Unilever in the Indian context, and Nestle and many others in the global context, are looking at the impact their operations are having on the environment as well as the positive or negative impact on the communities in which they operate. Understanding these impacts can drive improvements in corporate strategy, day to day operations and ultimately in financial performance.

4.1 Valuation of Ecosystem services: two cases

In this section, I talk of two cases of valuation of ecosystem services: one at the scale of an ecosystem, and the other at the scale of a landscape. Both the cases are studies conducted under the aegis of the *Ecological Economics programme* housed in the Policy, Research, and Innovation Division of World Wide Fund for Nature, India (Ghosh et al. 2015). The ecosystem chosen is that of the Kunigal wetlands in the southern state of India, Karnataka. Kunigal wetland is a peri-urban wetland located between 13.02 °N and 77.03 °E at an average elevation of 773 m in Kunigal, Tumkur district of Karnataka state. The lake has a total spread area of 416.20 ha and a gross water storage capacity of 532.2 MCFT. The lake is mainly fed by the Hemavathi River and from the rain water from the catchment area of 33,914 ha. The catchment mainly includes agricultural land, barren land, and scrubland.

On the other hand, the landscape chosen is the Terai Arc Landscape (TAL). TAL in Uttarakhand is among the most densely populated rural areas in the country as more than 8 million people reside there (2011 census). During the last two decades, the population in TAL has increased by as much as 54.2%, which is 9% above the national average. Most of the poorer communities in TAL depend on the forest for their subsistence.

4.1.1 Valuing ecosystem services of Kunnigal Wetlands

Wetlands generally provide important ecosystem services, upon which societies depend. They play a vital role in: contributing to food security by enabling direct availability of products such as fish, crops grown, wild fruits and vegetables; providing cash income from sale of raw materials and processed products; and contributing to increased crop and livestock yields as a result of improved productivity from use of water, silt, and through climate moderation. Apart from the various provisioning services, they also support various regulating (carbon sequestration, microclimate regulation) and cultural services (tourism).

The Kunnigal Lake provides many ecosystem services that include providing habitat for fish-breeding, water for domestic and agriculture, support for local biodiversity including migratory birds, aquatic vegetation, flood control, purification of wastewater, and groundwater recharge, among other services that are of immense value to local communities. The lake is an important religious place, with a shrine dedicated to the local deity—Someshwara temple.

The biodiversity of the lake includes 63 species of resident and migratory species of birds such as shovelers, pintail, pochard, spoonbill, painted stark, spot billed pelican, sandpiper, herons, whistling ducks, gadwal, and teals among others. Similarly the fish species included Common Carp, Catla, Rohu, Mrigal, Silver Carp, Grass Carp. Tilapia and Cat fish. The local people have been using the lake resources since ancient times to sustain their livelihoods. Currently the lake is leased to fishing co-operative to culture and harvest commercial fish.

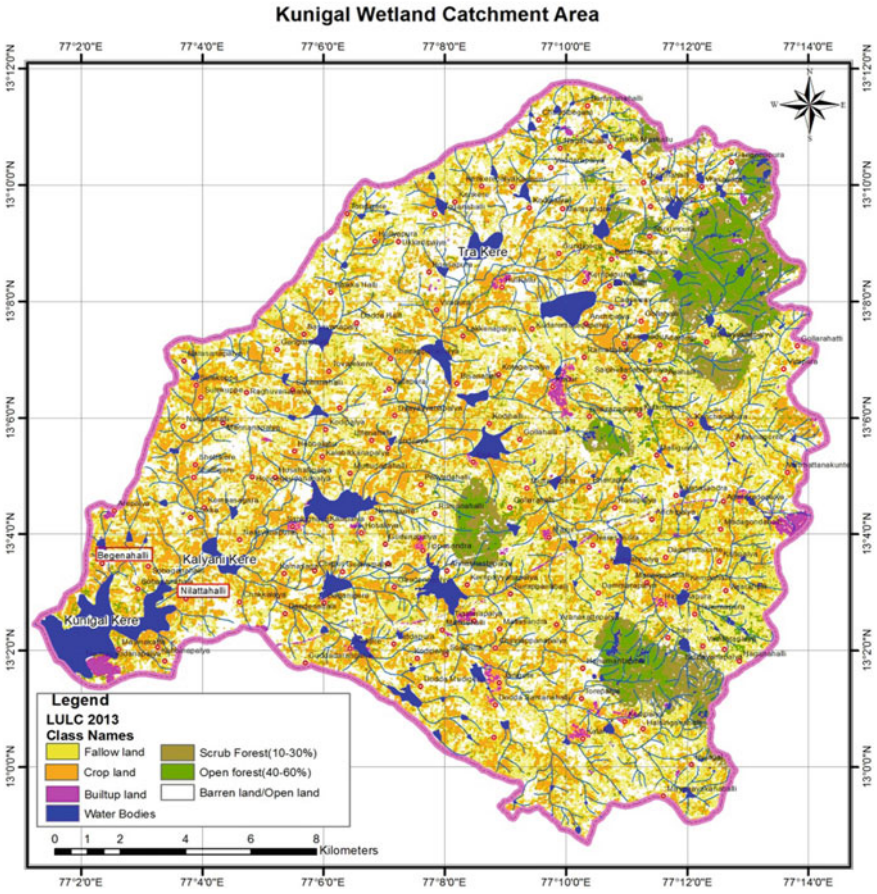


Fig. 3 Kunigal wetland catchment area

The lake is surrounded by a number of villages and the town of Kunigal (Fig. 3). For the current study, three villages–Bagenahalli, Neelathalli and Mavanakatte playa, located in the vicinity of the lake were selected. According to the census (2011) the total population of these villages were 1903 (920 males and 983 females) in 530 households. The villages have a primary school and there are no high schools, secondary schools or college.

A vast majority of the people living adjacent to the lake areas directly use the wetland resources for their sustenance. The valuation of Kunigal Lake in Tumkur is carried out by considering essentially seven ecosystem services, namely, water for agriculture, domestic water supply, fishery, and fodder, as provisioning services on one hand; and water purification, carbon sequestration, and micro-climate regulation as regulating services on the other. This valuation study has also gone ahead to

Table 1 Valuation of Selected Ecosystem Services of Kunigal Wetlands

Ecosystem Service	Value (in million)	Classification	% of each	Level
Domestic water use	25.56	Provisioning	2.96	Meso
Water for agriculture	11.8	Provisioning	1.37	Local/ Micro
Fishery	8.6	Provisioning	1.00	Local/ Micro
Fodder	1.4	Provisioning	0.16	Local/ Micro
Water purification	81.21	Regulating	9.41	Meso
Carbon Sequestration	749.26	Regulating	67.41	Global
Micro-Climate Regulation	152.61	Regulating	17.69	Meso
Total Value of Existing Ecosystem Services	**1030.45**			
Potential Tourism benefit	159.37 (with 137.26 million as potential revenue)	Cultural		

estimate the potential benefit annually if tourism is developed, as tourism is not yet developed in that region.

The summary findings of the study are given in Table 1.

Therefore, as shown in Table 1, the total value of ecosystem services in 2015–16 was INR 1030.45 million, while there was also an untapped potential of total tourism benefit of INR 159.37 million that can yield an annual revenue of INR 137.26 million.

In this context, it was found that the wetlands provide 24% higher value than the average incomes of the households. This creates a clear case for conservation, as the community at the local and meso levels will lose out 24% more than their annual incomes if the lake is lost. As such the loss will be even more if one considers the global benefit of carbon sequestration. This creates a clear case of conservation.

This study contributes to tackling under-investment in environmental assets through better economic analysis for environmental investments, including mobilization of government and donor resources for environment. In particular, it provides lessons for sustainably managing environmental resources to benefit local community which uses the wetlands as their main source of livelihood as well as improving its management for its sustainable use.

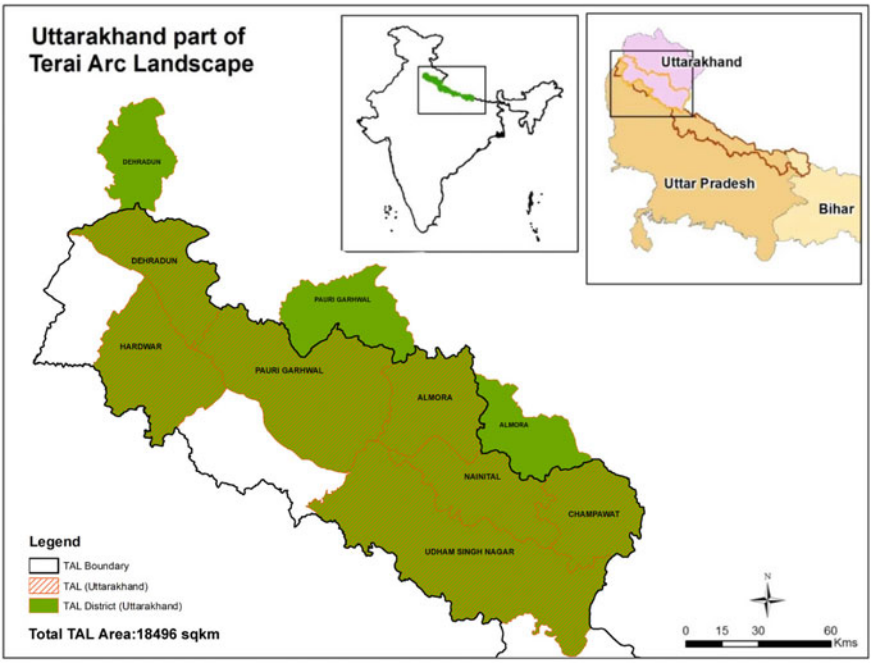


Fig. 4 Uttarakhand part of Terai Arc Landscape

4.1.2 Terai Arc Landscape in Uttarakhand

In the Terai Arc Landscape (TAL) in Uttarakhand, natural resource based occupations are predominant. Only 7% of the population uses purchased fuel such as liquefied petroleum gas (LPG), coal and kerosene in the entire TAL-India, the remainder use firewood collected from the forests. One needs to note here that the entire districts of Dehradun, Hardwar, Pauri Garhwal, and Almora do not fall in the TAL region (Fig. 4).

Values of nine ecosystem services of the TAL have been estimated. These are provisioning services like, water (used for agriculture, hydropower, and drinking water), fuel wood, and fodder, regulating services like carbon sequestration, and microclimate regulation, and cultural services like tourism (nature and pilgrimage). These have been estimated in the 2005-06 prices using standard valuation methods. Then the values of 2010–11 and 2015–16 have been arrived at by using Wholesale Price Index (WPI) as deflators.

The values of the ecosystem services in the landscape are given in Table 2.

The sum of the values of the nine ecosystem services in 2005–06 was INR 227.52 billion (US$3.5 billion). The same value turned out to be INR 390 billion (US$6 billion) in 2015–16, and INR 344 billion (US$5.3 billion) in 2011–12.

Table 2 Value of selected ecosystem services in Terai Arc Landscape in Uttarakhand

Item	Value in 2005–06 (Rs. million)
Water for agriculture	13886.82
Water for hydropower	440.68
Carbon sequestration	66078.20
Tourism (Corbett)	3680.00
Drinking water	2785.64
Fuelwood	41995.50
Microclimate regulation	48011.40
Fodder	3015.54
Religious tourism in hardwar	47623.51
Total	**227,517.28**

These are conservative estimates for various reasons. First of all, we have confined our analysis to a few selected ecosystem services, and have not extended it to obtain the full gamut of the services provided by the ecosystem. We have considered only nine ecosystem benefits and not the other ecosystem services like climate regulation, flood control, and many other services that Verma et al. (2015) considered while calculating economic valuation of select Tiger Reserves in India. This was more so because of the heterogeneity of the ecosystem that would have made estimation a complex affair. Secondly, there remains the problem of double counting while considering the supporting services of the ecosystem. To remove that possibility completely, we have not considered any supporting service in the analysis. Thirdly, while we have considered only certain aspects of the cultural services, e.g., religious tourism has been considered only for Hardwar, and nature tourism has been considered only for Corbett Tiger Reserve and the buffer and adjoining areas where the tourism has developed. There are many other aspects of nature tourism, e.g. Nainital, by itself, is a major tourist destination. Fourthly, the value of the benefits obtained by communities downstream of the landscape has not been considered. Carbon markets have been taken at one of the lows at USD 10 per ton of CO_2.

Next comes the question of ecosystem dependency of the community. In 2005–06, the estimated total income of the TAL districts was INR 191 billion. This is based on the estimates of population given by Census 2001, and district per capita income estimates by the Central Statistical Organisation. Therefore, the nine ecosystem services (estimated as INR 227.52 billion) yielded 19% more value than the total income of the region. It can therefore be argued that if the landscape ecosystem is destroyed through land use change, one needs to compensate the local community in TAL Uttarakhand by spending 19% more than the total value of the economic output of the landscape.

4.1.3 GDP of the poor

This paper has now discussed the importance and application of valuation of ecosystem services beginning from the ecosystem to the landscape scale. The values that we have arrived at are approximations, conservative, and "tip of iceberg" estimates. Yet, they are indicative of the fact that the contribution of the TAL ecosystem in Uttarakhand to the human community is at least INR 390 billion or USD 6 billion, which by itself is higher than the income of the community of the region. Similarly, a small ecosystem like the Kunnigal lake provides as much as USD 17 million.

It is clear that the population as a whole derives a substantial value from the ecosystem, which, at times, may be more than what they obtain from the economy. It is in this context, we also bring in the idea of the "GDP of the poor", as ecosystem services have been defined in the framework of TEEB (2010). The poor are more dependent on the ecosystem services than the rich (Martinez-Alier 2011).

In Kunnigal, the ecosystem dependency ratio (ratio of sum of values of ecosystem services and total incomes of the economy of the ecosystem) has turned out to be 1,24, while in 2005–06 in TAL, it was 1.19. However, one needs to note that as per our estimates, the ecosystem dependency ratio has been diminishing in TAL, as estimated in 2011 and 2015. This is because the incomes from other sources especially the tertiary sources that are not dependent on these nine ecosystem services have increased. The *ED*-ratio in 2011 is 0.52, and in 2015, it is 0.41. In any case, one may not ignore the fact the ecosystem services add up 40% more benefits to the earnings of the local community.

One needs to consider here that the ecosystem dependency ratio of the poor will be substantially higher than this. More than half the population in the TAL—Uttarakhand is living below poverty levels and an earning member of a household earns as little as US$ 1.9/day. The ecosystem dependency of these households is higher than those earning average per capita incomes. Hence, any policy towards land-use change in the landscape and ground actions leading to land use change in the wildlife habitats and corridors should be considered very carefully. One needs to take into account the scarcity value of the ecosystem services, i.e. the economic value loss with ecosystem service loss, as it is the poor who suffer the most from the loss in ecosystem services.

While land use change is planned, one needs to assess on how much of the habitat will be destroyed due to that. In those cases, poorer people will lose out a substantial amount of their "GDP" or "incomes" provided by the ecosystem, and they need to be compensated for the loss to the tune of the damage caused to them. However, this compensation would not take into account other economic impacts from the loss of services such as flood control, water recharge, and soil retention, which could lead to huge costs due to disasters incurred such as floods and landslides.

It therefore becomes important that the results of such an analysis are shared with policy makers to demonstrate that, in terms of economic development, it is critical to take into account the net cost of losing ecosystem services and the impacts of this on the rural poor. If a valuation from a long term development perspective is

undertaken, then it will be clear that the cost of damaging ecosystems and disrupting their services will be higher than the short term gains from some planned projects.

5 The Steps Forward …

In the wake of globalization and global standards of market efficiency, regulation, and competition, the Indian corporate sector needs to balance increased regulation, protecting the brand and ensuring stable supply chains with seeking opportunity for enhanced performance, thereby using the sustainability agenda for strategic advantage. The new strategic vision entails developing and integrating a detailed sustainability vision into the long-term strategic plan in a way that creates lasting value whilst also building public trust is a common challenge for all types of organisations.

Therefore, it becomes increasingly important for the firm to look at the values of the services provided by the ecosystem in which they plan to intervene. The footprint of the firm may be visible through its activity chain, beginning from procurement to marketing. As shown in the two cases of the valuations, firm needs to take into consideration this value for their own strategic decision making for a project in an ecosystem or the landscape.

At the same time, merely looking at one-year value may not be sufficient. The values of the short-run benefits, that the businesses often look at for justifying their projects, might be extinct in the medium and long-run. What might remain after the short run are apparently the costs incurred from lost ecosystem services that affect the broader human endeavours including livelihoods and businesses! Again, this impact might not only be a temporal phenomenon, but spatial too. While the intervention of the business might affect the ecosystem functioning, and eventually the services, the impacts of their operations might not be confined locally. It may extend much beyond through a ripple effect. Unfortunately, even linear infrastructure projects in India do not often take into consideration the value losses due to ecosystem service losses across space and time.

More importantly, *Sustainability Reports* are becoming popular these days. These reports consist of various initiatives on water management, resource management, energy management, employee benefits, community development, etc by the organisation, and are disseminated through various modes of communication. *Sustainability Reports* have become important strategic tools for positioning the organisation by informing stakeholders about the sustainability performance, and progress made over time in terms of the various indicators. To make sure how the organisation is accurately reporting on its corporate activities to support climate change, resource scarcity and socially responsible investing, assurance from a trusted adviser is sought. Valuation of ecosystem services will also be important here to report on not only the footprint, but also to report on the value created through afforestation measures that have been conducted by many companies.

What is becoming increasingly clear is that a sound sustainability strategy not only protects but also rebuilds a company's reputation, drives innovation and employee

engagement, it satisfies consumers and attracts and retains top talent. This is akin to CSV in the way conceived by Porter and Kramer (2011). At a different level, it demonstrates compliance and leads to market differentiation, thereby helping positioning of the organisation with a unique reputation in the market. In this context, integrating ecosystem services values in firm's decision making for better ex ante or ex post impact assessments, and setting conservation priorities needs to be seen as a cornerstone of sustainability.

To conclude, the values of ecosystem service losses/gains need to be taken into consideration in the benefit-cost ratio of any project intervening into the working of the ecosystem (Barber et al. 2014). These values can do two things: a> if the net benefit-cost ratio is greater than unity, then it justifies that the firm's interventions are not only financially viable for itself, but also socially and ecologically viable; b> if the benefit-cost ratio is less than unity, this might lead the firm to alter investment decisions. Such valuation can also help the firm in placing social, economic, and ecological justifications of their projects in the public forum. While embracing sustainability as corporate strategy is about creating sustainable bottom-lines with the firm's embedment in the broader ecosystem, valuing ecosystem services and integrating them in the decision-making is about creating shared value.

Acknowledgments This paper is largely based on the multiple research projects/papers of the author in his capacity as Senior Economic Adviser at WWF-India. Substantial parts of the ideas presented in this chapter were first published in 2016 by Policy Research and Innovation Division of WWF-India as Issue Brief No. 2 titled, "Valuing Ecosystem Services at the Scale of a Large Mammal Landscape: The Case of the Terai Arc Landscape in Uttarakhand", co-authored by Nilanjan Ghosh, Dipankar Ghose, G. Areendran, Divya Mehra, Ambica Paliwal, Krishna Raj, Kiran Rajasekariah, Ambika Sharma, Anil Kumar Singh, Shashank Srinivasan and Sejal Worah. The same has been reproduced here with permission from WWF India, which further attributed it under Creative Commons.

References

Barber, C. P., Cochrane, M. A., Souza, C. M., Jr., & Laurance, W. F. (2014). Roads, deforestation, and the mitigating effect of protected areas in the Amazon. *Biological Conservation, 177,* 203–209.

Bockstael, N., Freeman, A., Kopp, R., Portney, P., & Smith, V. (2000). On measuring economic values for nature. *Environmental Science and Technology, 34*(8), 1384–1389.

Chopra, K., & Adhikari, S. K. (2004). Environment development linkages: Modelling a wetland system for ecological and economic value. *Environment and Development Economics, 9*(1), 19–45.

Costanza, R., d'Arge, R., de Groot, R., Farberk, S., Grasso, M., Hannon, B. et al. (1997). The value of the world's ecosystem services and natural capital. *Nature*, *387*, 253–60.

Costanza, R., de Groot, R., Sutton, P., van der Ploeg, S., Anderson, S. J., Kubiszewski, I. (2014). Changes in the global value of ecosystem services. *Global Environmental Change*, *26*, 152–58.

Ghosh, N. (2015). *Need to make conservation everybody's business*, a. Retrieved October 14, 2017 from http://www.abplive.in/blog/need-to-make-conservation-everybodys-business-239921?rs_type=internal&rs_origin=section&rs_medium=article&rs_index=6.

Ghosh, N., Rajashekariah, K., Babu, S., & Gopala, A. (2015). *Valuation of ecosystem services of Kunigal Wetlands in Tumkur District, Karnataka, Unpublished draft for circulation*. WWF-India.

Ghosh, N., Tapsall, P., & Shunmugavel, A. (2016). *Sustainability as corporate strategy: The emerging dimension, PRI commentary No. 1*. WWF India.

Ghosh, N., Ghose, D., Areendran, G., Mehra, D., Paliwal, A., Raj, K., et al. (2016a). *Valuing ecosystem services at the scale of a large mammal landscape: The case of the Terai Arc landscape in Uttarakhand, policy research and innovation division, Issue Brief No. 2*. New Delhi: WWF-India.

Krehmeyer, D., Lenox, M., & Moriarty, B. (2010). *Sustainability must be central to corporate strategy now*. Retrieved October 14, 2017 from https://www.forbes.com/2010/06/16/sustainability-corporate-strategy-leadership-citizenship-ethisphere.html.

Martinez-Alier, J. (2011). Environmental justice and economic de-growth: An alliance between two movements. *Capitalism, Nature, Socialism, 23*(1), 51–73.

Meadows, D. H., Meadows, D. L., Randers, J., & Behrens, W. W., III (Eds.). (1972). *The limits to growth: A report for the club of Rome's project on the predicament of mankind*. New York: Universe Books.

Millennium Ecosystem Assessment (MA). (2005). *Ecosystems and human well-being: Synthesis*. Washington: Island Press.

Pearce, D. W., & Turner, R. K. (1989). *Economics of natural resources and the environment*. Johns Hopkins University Press.

Porter, M. E., & Kramer, M. R. (2006). Strategy and society: The link between competitive advantage and corporate social responsibility. *Harvard Business Review, 84,* 78–85.

Porter, M., & Kramer, M.R. (2011). Creating shared value: Redefining capitalism and the role of the corporation in society. *Harvard Business Review*, January–February 2011 Issue.

TEEB. (2010). *The economics of ecosystem and biodiversity: Mainstreaming the economics of nature: A synthesis of the approach, conclusions, and recommendations of TEEB*.

Verma, M., Negandhi, D., Khanna, C., Edgaonkar, A., David, A., Kadekodi, G. et al. (2015). *Economic valuation of tiger reserves in India: A Value+ Approach*. Bhopal, India: Indian Institute of Forest Management, January 2015.

WBCSD. (2008). *Measuring Impact Framework*. Geneva: World Business Council for Sustainable Development.

WBCSD. (2013). *Eco4Biz: Ecosystem services and biodiversity tools to support business decision-making Version 1*. Geneva: World Business Council for Sustainable Development.

Insider Ownership and the Performance of Firms in India: Evidence from a Panel Semi-parametric Regression Model

Indrani Chakraborty

1 Introduction

There has been an intense debate in corporate finance literature on the issue of relationship between ownership and firm performance. This debate goes back to the thesis of Berle and Means (1932) which argued to have found an inverse relationship between dispersed shareholders and firm performance. According to this view, if the ownership structure of a firm is dispersed in the hands of a large number of minority shareholders they are unable to exert much influence on the management of the firm, and the managers overwhelmingly control the affairs of the firm. The managers in this case would tend to use corporate assets for their own benefits instead of maximizing shareholders' wealth. The suggested solution to this agency problem is known to be giving managers an equity stake in the firm, what is called 'insider ownership'. This helps resolve the moral hazard problem by aligning managerial interests with shareholders' interests (Jensen and Meckling 1976; Fama 1980; Jensen and Murphy 1990; Jensen 2000).

There are two competing hypotheses in the existing theoretical literature regarding the effects of insider ownership on firm performance, viz. the convergence-of-interests hypothesis and the entrenchment hypothesis. The first one argues that an increasing insider ownership aligns the manager's interests with outside shareholders' and hence results in a positive effect on firm performance. Wruck (1988) and Mehran (1995) provided empirical support in favour of this hypothesis. On the contrary, the entrenchment hypothesis suggests a negative effect of insider ownership on firm performance, since higher insider managerial shareholdings are likely to shelter insiders from the influence of the market for corporate control (Fama and Jensen 1983).

I. Chakraborty (✉)
Institute of Development Studies Kolkata, Bidhan Nagar, India
e-mail: indrani.c61@gmail.com

P. Ray et al. (eds.), *Economics, Management and Sustainability*,
https://doi.org/10.1007/978-981-13-1894-8_11

In an attempt to synthesize the two rival arguments, a growing body of research has suggested the existence of nonlinear relationships between insider ownership and firm performance. However, the empirical findings so far have been rather diverse because of the inherent complexity of any presumed nonlinear model. For example, Morck et al. (1988) presented an N-shaped curve with two inflexion points; Hermalin and Weisbach (1991) depicted the relationship as a M-shaped curve with 3 inflexion points; Cui and Mak (2002) found a W-shaped relationship with 3 inflexion points; Davis et al. (2005) specified a fifth-degree function with two maximum inflexion points and two minimum inflexion points; Selarka (2005) found the relationship as a U-shaped curve with one inflexion point. However, the main limitation of all these studies is that the relationship between insider ownership and firm performance has been specified in an arbitrary manner. An alternative method is to estimate a panel semi-parametric regression model, the advantage of which is that it is a data-driven estimation method where we do not specify and impose any a priori functional form restriction on the data on insider ownership and firm performance. In this regression method, firm performance will be included nonparametrically in the regression model, whereas all other control variables will be included linearly. The true functional relationship between insider ownership and firm performance will be revealed from the data itself.

The objective of this study is to examine the relationship between insider ownership and firm performance in Indian listed firms. To examine this issue we apply the panel semi-parametric regression technique. As far as our knowledge goes, in the empirical research on the relationship between insider ownership and firm performance, the panel semi-parametric regression technique has not been applied so far. In this respect, the study will have some methodological contribution to make. Another important issue that has been addressed by several researchers in this context is the issue of endogeneity (Demsetz 1983; Demsetz and Lehn 1985). However, some recent papers show that the issue of endogeneity can be taken care of by using the generalized method of moments (GMM) estimation (Wintoki et al. 2009). As a robustness test, in our study, we apply system dynamic panel estimation technique, which is based on GMM method, controlling for the problem of endogeneity.

The second contribution of this study is to propose a new structure to the insider ownership–performance relationship which captures a more complex characterization of the evolving behaviour of managers in Indian firms dominated by business groups.[1] We argue that with equity holdings above 50%, although the managers have substantial control of the firm, the internal governance resulting from the corporate governance mechanism will lead to convergence of interests. Only at very high levels of managerial holdings, above 80% of equity holding, the entrenchment effect predominates.

[1] Business groups are companies owned by a business family. Group-affiliated firms consist of groups of companies that are connected through a network of legal, financial and transactional relationships. Indian business groups are characterized by a substantial amount of intra-group financing (Gopalan et al. 2007).

Applying panel semi-parametric regression technique to a large sample of Indian firms, insider ownership is seen to have a significant impact on firm performance and the observed relationship is supportive of our proposed structure of a quartic relationship. The dynamic panel estimation method, taking care of the problem of endogeneity, also supports this relationship.

The rest of the paper is organized as follows: Sect. 2 provides the analytical framework. Section 3 gives a brief literature review. Section 4 discusses the methodology. Section 5 explains the data and variables. Section 6 presents the empirical results, and Sect. 7 concludes.

2 Analytical Framework[2]

We propose an alternative structure to the relationship between insider ownership and firm performance and argue that, in the Indian context, where the dominance of business groups in the corporate structure is prevalent, the cubic, or piecewise linear, representation used in earlier studies is inappropriate. The model that is represented here captures further nonlinearities in this relationship at high levels of insider ownership and has a quartic specification. There are both positive and negative incentives for the insiders to maximize firm value, and these incentives operate differently at different levels of insider ownership. In this context, three factors play important roles, viz. convergence of interest, entrenchment and the role of corporate governance.

In group-affiliated firms family members play a larger role in the management of the group than in the case of stand-alone firms (Holderness and Sheehan 1988; Denis and Denis 1994). The involvement of members of the owner family may further contribute to the alignment of interest between owners and managers (Lemmons and Lins 2003). Furthermore, the active involvement of family members in group-affiliated firms may be beneficial in terms of the long time horizon with respect to which investment and other strategic decisions are taken (James 1999). Moreover, due to the longer time horizon of the family members they are more likely to cooperate and make decisions that maximize firm value (Anderson and Reeb 2003). Thus, for initial increases in managerial ownership up to a level of 20–25%, this argument holds good and, hence, firm performance increases.

The greater family involvement in the management of group-affiliated firms as compared to stand-alone firms is likely to exacerbate entrenchment effects (Morck et al. 2005). Moreover, similar to stand-alone firms, group-affiliated firms hire managers from outside the family, too. In the absence of familial ties, agency problems with non-family managers are more likely because emotional and psychological bases of reciprocal altruism tend to be weaker (Chua et al. 2003). As the managerial ownership increases further around 50% level, the non-family manager may extract pecuniary benefits for herself or family or she may take decisions that favour ‘cronies’ or hire incompetent relatives for key posi-

[2]This section is partly drawn from Richter and Chakraborty (2015).

tions (Bloom and Van Reenen 2007). Moreover, information asymmetry between the dominant managers and minority shareholders may increase the entrenchment effect due to lower flow of information. Less transparency will affect performance adversely (Wang 2006). Unlike in USA, the poorly performing managers may not be removed by hostile takeovers in India. Mergers require three-fourths majority of shareholders under Indian company law (Mathew 2007). Thus, even owning a little over 25% of share the managers can effectively block takeover. However, if a manager is from within a family, this situation may entail other problems. For example, family managers may be less competent than outsiders as they come from a smaller selection pool (Volpin 2002) and hence lead to a decrease in firm value.

At above 50% ownership, the manager will have complete control of the company, but will be subject to some internal governance. In group-affiliated firms a particular kind of agency problem arises due to the conflicts between controlling shareholders and minority shareholders. Controlling shareholders may expropriate minority shareholders to extract private benefits for themselves (Morck and Yeung 2004). One way that controlling shareholders expropriate minority shareholder wealth is by tunnelling (Shleifer and Vishny 1997). The pyramidal structures of group-affiliated firms, along with internal markets for capital and labour and related party transactions, may facilitate the expropriation of minority shareholders through distribution of group profits across affiliates. These expropriation practices of the controlling shareholders over minority shareholders in group-affiliated firms may ultimately reduce profitability (DeAngelo and DeAngelo 2000; Morck et al. 2000). However, in the Indian context, the minority shareholders will put pressure on the controlling shareholders to act in the best interests of the shareholders so that value maximization gets priority and firm performance increases. Due to the specific characteristics of corporate governance in India which emphasize the importance of independent directors in the board, the manager will be under pressure to maximize firm value. These independent directors protect the interests of the minority shareholders in Indian firms. After liberalization of the Indian economy, serious efforts have been made towards overhauling the system of corporate governance. Clause 49 of the Listing Agreements to the Indian stock exchange comes into effect from 31 December 2005, which was formulated for the improvement of corporate governance (Chakrabarti et al. 2007). Firms that do not comply with Clause 49 can be de-listed and charged with financial penalties. The constitution and functioning of the board of directors emerges as the key area of focus for Clause 49. It requires that the non-executive members should comprise at least half of a board of directors. It further requires that the board of directors must have a minimum number of independent directors. If the chairman is an executive director, at least one half of the board should comprise independent directors. In other cases, independent directors should constitute at least one-third of the board size. Therefore, at the managerial ownership level of above 50%, the convergence of interests plays a role here which leads to increase in firm value.

Finally, as managers' share exceeds 80%, the manager becomes the sole owner of the company and no external discipline or internal control will be effective to force the manager towards value maximization, leading to entrenchment. Tunnelling behaviour of the group-affiliated firms becomes operative at this level of managerial

ownership, evidence of which is provided by Bertrand et al. (2002) in the particular context of Indian firms.

3 Review of Earlier Literature

The impact of insider ownership on firm performance has been a topic pursued in a large number of studies. The first studies investigating this relationship have been those of Morck et al. (1988) and McConnell and Servaes (1990), where insider ownership is measured by top managements' shareholdings. Both studies found a significant nonlinear relationship. They show that as the ownership increases, firm value increases, but when ownership is too concentrated, the value of the firm starts decreasing. The inflexion points found in these two studies are different. Morck et al. (1988) using a sample of 371 firms in 1980 and using piecewise linear regression observe a positive relationship between managerial ownership and Tobin's q in the 0–5% range and beyond the 25% level, while as ownership increases from 5 to 25%, Tobin's q decreases. On the other hand, McConnell and Servaes (1990) perform a cross-sectional regression of firms in two separate years, 1976 and 1986, and found an inverse U-shaped relationship between managerial ownership and firm value. The relationship is maximized for the managerial ownership in the range between 40 and 50%. Therefore, these two studies support the existence of both the alignment effect and entrenchment effect. Short and Keasey (1999) study the relationship between ownership and firm value in the UK. They argue that there are important differences in corporate governance systems of the UK and the USA and hence the alignment effect and entrenchment effect will operate at different levels of managerial shareholding. They observed a cubic relationship between managerial ownership and firm value based on a sample of 225 UK listed firms over the period from 1988 to 1992. From their findings they conclude that UK managers become entrenched at higher levels of ownership than the US managers because of differences in corporate governance mechanism of the two countries. Faccio and Lasfer (1999) using a cubic specification report similar findings for high-growth firms in the UK. Mudambi and Nicosia (1998) examine the impact of ownership and control on performance of UK financial services for the years 1992–1994. They found that the increased control by largest shareholding group is positively related to firm performance, whereas increased concentration is inversely related to performance. In the Indian context, Sarkar and Sarkar (2000) study the nonlinear relationship between managerial ownership and firm value using piecewise linear regression method and they found firm value declined till 25% of managerial ownership and increased thereafter. Another important study in the Indian context is Khanna and Palepu (1999) which found insider ownership has a positive and significant impact on firm value, while director's holding has no impact.

A large number of recent studies have reported a nonlinear relationship of various shapes between managerial ownership and firm value. Hermalin and Weisbach (1991), based on a panel data of 142 NYSE firms observed over different years

between 1971 and 1983, and using piecewise regression, found an M-shaped relation between managerial ownership and firm performance with three inflexion points, viz. at 1, 5 and 20%. Cui and Mak (2002) examine the relationship using a sample of 3100 high R&D firms that are listed on NYSE, AMEX and NASDAQ in 1996 and 1998. Using piecewise regression analysis the study reveals that the relationship could be described as a W-shaped relationship. It was found that Tobin's Q initially declines as managerial ownership increases from 0 to 10%, increases between 10 and 30% and declines again between 30 and 50% and finally Q increases above 50% ownership. Thus, the findings of this study differ from that of Morck et al. (1988) and McConnell and Servaes (1990). Using data on 802 UK companies for the year 1995, Davies et al. (2005) argue that in the relationship between Tobin' s q and managerial ownership there are two maximum and two minimum inflexion points. Hence, they estimate a quintic specification between managerial ownership and firm value and obtain statistically significant results. The turning points occur at 7.9, 26.5, 55.2 and 86.2%. Thus, the findings of this study differ from the other UK studies by Short and Keasey (1999) and Faccio and Lasfer (1999) who report two inflexion points. Selarka (2005) using a sample of 1397 Indian firms for 2001 found a U-shaped relationship between managerial ownership and firm value with an inflexion point at 45.11%. She also reports significant and high difference between the inflexion point across group and stand-alone firms in the sample. The findings show that stand-alone firms have a very high point of inflexion, ranging from 51 to 75%. Hung and Chen (2009) found a V-shaped specification with a single threshold value at 5.4%, for a panel of data of 62 Taiwanese SMEs for the period from 1999 to 2006. Hence, the observed threshold in this study is much lower than Morck et al.'s (1988) value of 25%. Therefore, all these studies confirm that the rival effects of alignment of interests and entrenchment exist simultaneously, which lead to a nonlinear relationship between insider ownership and firm value. But it is evident that the nature of the nonlinear relationship varies from one study to another.

However, most of these studies have assumed ownership as exogenous and have been criticized by Demsetz (1983) and Demsetz and Lehn (1985). The latter studies argue that insider ownership is an endogenously determined variable and any observed correlation between ownership and firm value is meaningless. They argue that the relationship between insider ownership and firm value might be due to some firm-specific variables that are unobservable. Demsetz and Lehn (1985) show that ownership structure of US firms is determined by firm size, industry affiliation and some other firm-specific variables which support that ownership is endogenously determined. In a later study, Himmelberg et al. (1999) use a fixed effect panel data method and instrumental variables to control for unobserved firm-level heterogeneity. They found that the managerial ownership has no statistically significant effect on firm performance. On the other hand, assuming endogeneity of managerial ownership and applying simultaneous equation framework some other studies observe reverse causality (Cho 1998; Loderer and Martin 1997; Kole 1996). As the two studies referred to earlier, viz. Hermalin and Weisbach (1991) and Hung and Chen (2009), use panel data method, which allows to control for possible biases due to the joint endogeneity of managerial ownership and firm value, they may be considered

as an important improvement on the remaining studies in the present context. Both of these studies noted a nonlinear relationship between managerial ownership and firm value, as discussed above.

4 Methodology

To study the relationship between firm performance and insider ownership the most commonly used parametric specification is as follows:

$$\text{Perf}_{it} = \alpha_i + \beta_1 \text{OWNINS}_{it} + \beta_2 Z_{it} + \varepsilon_{it} \ldots \quad (1)$$

where Perf_{it} is the measure of firm performance, for example Tobin's Q, in firm i in year t, OWNINS_{it} is the insider ownership in firm i in year t, Z_{it} represents other control variables in firm i in year t, and α_i is the firm fixed effects. As discussed earlier, in the existing literature the relationship between Perf_{it} and OWNINS_{it} in Eq. (1) is found to be nonlinear. To capture the nonlinearity between Perf_{it} and OWNINS_{it}, in the earlier studies either quadratic or cubic terms on OWNINS_{it} are included arbitrarily. But, it is well known that misspecification of the functional form can lead to biased estimation and hypothesis testing. Hence, we are applying the panel semi-parametric method of estimation of Eq. (1) where the control variables are included linearly and the variable OWNINS_{it} appears nonparametrically. In this method, we do not impose the functional form a priori and hence we can avoid the problem of misspecification of the functional form. We are using Baltagi and Li (2002) panel semi-parametric estimation method with fixed effects. The details of the approach are as follows. Let $Y_{it} = \text{Perf}_{it}$ and $X_{it} = \text{OWNINS}_{it}$, then we can specify a partially linear panel data model as follows:

$$Y_{it} = m(X_{it}) + Z_{it}\beta + \alpha_i + \varepsilon_{it} \ldots \quad (2)$$

where m(.) denotes the unknown (nonparametric) functional form. To remove the fixed effects, following the first differencing method for linear panel data models, we consider the following first-order differencing model:

$$Y_{it} - Y_{it-1} = m(X_{it}) - m(X_{it-1}) + (Z_{it} - Z_{it-1})\beta + \varepsilon_{it} - \varepsilon_{it-1} \ldots \quad (3)$$

In Eq. (3), Baltagi and Li (2002) proposed to approximate $m(X_{it}) - m(X_{it-1})$ by the following series differences:

$$p^k(X_{it,} X_{it-1}) = \left[p^k(X_{it}) - p^k(X_{it-1})\right] \ldots \quad (4)$$

where $p^k(X)$ are the first k terms of a sequence of functions $(p_1(x), p_2(x), \ldots)$.

Thus, (2) becomes as follows:

$$Y_{it} - Y_{it-1} = \left(p^k(X_{it}) - p^k(X_{it-1})\right)\theta + Z_{it} - Z_{it-1})\beta + \varepsilon_{it} - \varepsilon_{it-1} \ldots \quad (5)$$

which can be estimated consistently using OLS. Having estimated Θ and β, it is easy to fit the fixed effects α_i. Then one has to go back to Eq. (2) to estimate the error component residual:

$$u_{it} = Y_{it} - Z_{it}\beta - \alpha_i = m(X_{it}) + \varepsilon_{it} \ldots \quad (6)$$

Equation (6) is a nonparametric panel data model, where the curve m(.) can be fitted by regressing u_{it} on Z_{it} using some standard nonparametric regression estimators such as spline regression and local linear smoothing techniques. In our empirical analysis, to estimate m(.) we apply B-splines which is a fractional polynomial with pieces defined by a sequence of knots $c_1 < c_2 < \cdots < c_k$, where they join smoothly. B-splines, for a set of $k+2$ consecutive knots, are defined as follows:

$$B(Z,\ c_1 \ldots c_{k+2}) = (k+1)\sum_{j=1}^{k+2}\left[\prod\nolimits_{1\le h\le k+2, h\ne j}\left(c_h - c_j\right)\right]^{-1}\left(z - c_j\right)_+^k$$

B-splines are a rescaling of each of the piecewise functions. For further details on B-spline see Newson (2001).

5 Data and Variables[3]

5.1 Data

The sample for India is drawn from PROWESS, a database provided by Centre for Monitoring Indian Economy (CMIE). The sample was chosen for all Indian firms listed in the Bombay Stock Exchange (BSE) and National Stock Exchange (NSE) for the period 2006–2013. We begin our analysis from 2006 as Clause 49 of the Listing Agreements to the Indian stock exchange came into effect from 31 December 2005, which was formulated for the improvement of corporate governance. We, therefore, expect that better reporting on corporate governance will be followed by Indian firms since 2006. We have eliminated those firms for which information on shareholding patterns and other variables is missing. After excluding firms on the above basis, a final sample of 368 firms with 3937 observations is derived.

[3]This section draws from the author's previously published work titled 'Effects of ownership structure on capital structure of Indian listed firms: Role of business groups vis-à-vis stand-alone firms', published in Emerging Markets Finance and Trade, 2018.

Table 1 Pattern of insider ownership over the years in sample firms

Insider ownership (%)	2006	2007	2008	2009	2010	2011	2012	2013
Mean	50.02	49.88	50.13	51.19	51.24	51.38	51.90	52.22
Std. dev.	15.27	15.17	15.06	14.95	15.28	15.44	15.40	15.36
Minimum	8.78	8.73	8.85	5.30	5.30	5.30	5.30	5.12
Maximum	98.19	98.19	98.19	98.19	98.19	98.19	98.19	98.19

5.2 Variables

5.2.1 Performance Measures

As the measure of firm performance we are using a market-based measure, viz. Tobin's q. Tobin's q is measured by the ratio of sum total of market value of equity and book value of debt to total asset.

5.2.2 Measure of Insider Ownership

The variable of interest for the present study is the amount of share ownership by insiders (OWNINS). This is measured as the percentage of shares owned by the promoters,[4] which has been used in earlier studies on Indian firms as the measure of insider ownership (Sarkar and Sarkar 2000; Pattanayak 2001). Data relating to the insider ownership for the period 2006–2013 for India are shown in Table 1. The evidence suggests that insiders hold approximately 50% of the ownership rights in the firms contained in the sample and that this proportion increased slightly over time. The minimum percentage of equity holding by insiders decreased from 8.78% in 2006 to 5.12% in 2013, whereas the maximum percentage of their equity holding (98.19%) remained unchanged.

5.2.3 Control Variables

A number of other variables are included as control variables which affect firm performance. These include firm size (SIZE), firm growth (GROWTH), leverage (LEV), research and development expenditure (R&D) and age of the firm (AGE). Table 2 summarizes the description of all the variables used in this analysis.

[4]Promoters are defined as all individuals and their relatives, corporate bodies/trusts/partnership or any other type of entity that either founded or acquired a controlling stake in the firm concerned, where the ownership stakes exceeds that of any external shareholder.

Table 2 Summary table on description of the variables

Variable definition	Symbol used	Measurement of the variable
Tobin's q	Q	Ratio of sum total of market value of equity and book value of debt to total assets
Insider ownership	OWNINS	Percentage of equity shares owned by the promoters in Indian firms and percentage of equity shares owned by the directors and their immediate families for UK firms
Leverage	LEV	Ratio of total borrowing to assets
Size of firm	SIZE	Natural log of sales
Age of firm	AGE	Log of number of years since the incorporation of the firm
Firm growth	GROWTH	Percentage change in sales
R&D expenditures	R&D	Research and development expenditures to total assets

6 Empirical Findings

6.1 Results of Panel Semi-parametric Regression

The results of the estimation for panel semi-parametric regression are presented in Table 3. Columns (1) and (2) report the parametric estimation results. Column (1) reports the fixed effect linear panel model for Tobin's q. Column (2) reports the fixed effect cubic panel model for Tobin's q to examine the cubic specification as suggested by Morck et al. (1988) and others. In column (3), we report the estimates of the semi-parametric panel fixed effect model in which firm performance is a nonparametric function of insider ownership (OWNINS) and is a linear parametric function of all other control variables.

The results show that OWNINS has no significant effect on firm performance in col. (1) and OWNINS, $OWNINS^2$ and $OWNINS^3$ are not significant in col. (2). Thus, the fixed effect linear panel and fixed effect cubic panel models do not turn out to be good fit. The estimates of semi-parametric panel fixed effect model show the graphical plot of the relationship between Tobin's q and insider ownership (OWNINS) in Fig. 1, which reveals a quartic relation between the two. We find that the alignment effect is supported for the insider ownership level in the ranges 0–20% and 50–80%. On the other hand, the entrenchment effect is supported for the ownership level 20–50% and above 80%. Thus, the findings are consistent with our theoretical predictions. These results are in conformity with Faccio and Lasfer (1999) who report the first two inflexion points at 19.68 and 54.12% from UK sample. But it differs from the studies by Morck et al. (1988) (5 and 25%), Cho (1998) (7 and 38%) and Short and Keasey (1999) (12.99 and 41.99%) in respect of inflexion points. While these studies limited the inflexion points to two, our study reveals that there is another inflexion point at much higher levels of insider ownership. It appears that internal

Table 3 Estimation results for linear panel, cubic panel and semi-parametric panel regressions for Tobin's q for Indian firms

Variables	Tobin's q		
	(1)	(2)	(3)
OWNINS	2847.93 (3246.34)	−7201.43 (32722.05)	
$OWNINS^2$		153.83 (709.45)	
$OWNINS^3$		−0.62 (4.73)	
SIZE	−247714.8 (61214.18)*	−244866.4 (61463.4)*	−314419 (78982.68)*
GRTH	3870.11 (580.52)*	3870.21 (580.93)*	2161.20 (436.41)*
LEV	−1031865 (149900.1)*	−1033874 (150035.6)*	−1700421 (162753)*
R&D	14.85 (40.06)	14.99 (40.09)	22.72 (55.78)
AGE	−1231579 (250866.7)*	−1248474 (254301)*	−1871555 (454330.2)*
Constant	7413832 (668862.7)*	7630127 (871159)*	
No. of observations	2937	2937	2565
R^2: overall	0.006	0.007	
Adj. R^2: within			0.06

Note (i) Standard error is presented in parentheses
(ii) * implies significance at 1% level
(iii) ** implies significance at 5% level
(iv) *** implies significance at 10% level

governance, through the proactive role of the independent directors on the board, has an influence on managerial ownership up to the point where the board takes complete control (80%) and Tobin's q decreases thereafter.

The control variables SIZE, LEV and AGE are negative and significant at 1% level in all the three specifications. The variable GROWTH is positive and significant at 1% level in all the three specifications. However, R&D is positive but not significant in any of the specifications. Theoretically speaking, firm size (SIZE) is expected to have a positive effect on firm performance due to two reasons. First, the larger firms have internal capital market advantage as well as better access to external funds which lead these firms to make greater use of those projects which are profitable. Second, larger firms enjoy greater economies of scale which, in turn, increases their profitability (Short and Keasey 1999). Thus, our result contradicts those as expected. Leverage (LEV) is also expected to have a positive effect on firm performance. The argument is

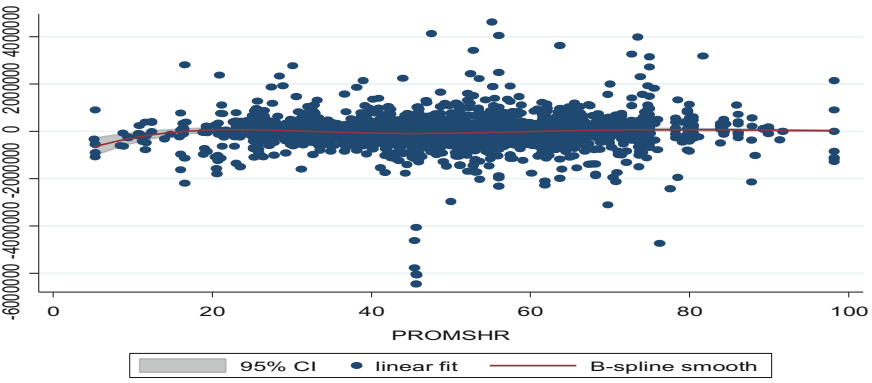


Fig. 1 Semi-parametric panel regression results for Tobin's q for Indian firms

that higher amount of debt in total assets may be used as a signal by the management that the firm has enough cash flows necessary to repay debt. Debt, therefore, may be used to resolve the agency conflict between manager and shareholders and hence would increase the value of the firm (Jensen and Meckling 1976). However, our result is conflicting with this argument. The reason may be that as leverage increases in Indian firms, bankruptcy threat for firms also increases which affects adversely the firm performance. Our result for the variable AGE supports that younger firms are more profitable than the older ones. This may happen due to the reasons that the older firms sometimes lack initiatives to diversify and cater the relevant consumer demand and sometimes they fail innovation leading to higher profitability (Pattanayak 2001). Then the result for firm growth (GROWTH) can be explained from the perspective of competition. The market share of a firm determines relative competitive position which has a positive impact on the firm performance. Firm growth measured as sales growth is an indicator of the competitive position and higher is the firm growth higher would be the firm performance. Therefore, the above-mentioned findings differ from some earlier studies in the context of India which include Sarkar and Sarkar (2000), Pattanayak (2001) and Chhibber and Majumdar (1999).

Thus, our findings do not provide support for the functional form of the relationship between performance of firms and insider ownership as suggested by some earlier studies including Morck et al. (1988) and others. Rather, we find a quartic relationship between the two from our semi-parametric panel fixed effect estimation. Both Morck et al. (1988) and Short and Keasey (1999) provided support for a cubic relationship between firm performance and insider ownership. However, the cubic specification has the limitation that inflexion points are sensitive to the inclusion of control variables. By dropping one or other control variables, the results get affected. Moreover, it may not fit the data well if the nonlinear relation is not smooth (Pattanayak 2001). This problem can be resolved by using spline regression as done by Morck et al. (1988). However, the inflexion points in spline regression are crucial. Morck et al. (1988) chose the inflexion points arbitrarily which might make the

Table 4 Piecewise panel regression results for Tobin's q for Indian firms

Variables	Tobin's q
promshr1	24583.85 (34853.78)
promshr2	−28983.79 (37504.59)
promshr3	15088.15*** (9830.38)
promshr4	−44842.48 (33040.65)
SIZE	−248229.5 (61412.59)*
GRTH	3851.43 (580.67)*
LEV	−1046847 (150284.2)*
R&D	15.62 (40.06)
AGE	−1258113 (253475.2)*
Constant	7268529 (921224.6)*
No. of observations	2937
R^2: overall	0.007

Note The same as Table 3

results biased due to misspecification of the model. To compare with Morck et al. (1988) we have also estimated piecewise linear regression, in a panel set-up, while identifying the inflexion points from our semi-parametric panel regression graphs.

Figure 1 shows three inflexion points at 20%, 50% and 80% which lead to four spline nodes at [0–0.20], [0.20–0.50], [0.50–0.80] and above. Based on these splines, Table 4 shows that for Tobin's q none of the splines is significant except the third spline which is significant at 10% level. Hence, piecewise panel regression results are not in conformity with our semi-parametric panel regression results. However, among the control variables, SIZE, LEV and AGE are negatively significant at 1% level, GROWTH is positively significant at 1% level, and R&D is positive but not significant, as shown in Table 3.

6.2 Results of Dynamic Panel Regression[5]

To test the robustness of our semi-parametric panel regression results we bring into consideration the role of endogeneity in the insider ownership–performance relationship as emphasized earlier by Demsetz (1983), Demsetz and Lehn (1985) and Himmelberg et al. (1999). These earlier literature emphasized on two sources of endogeneity that may bias the estimated relationship between insider ownership and performance, viz. unobservable heterogeneity and simultaneity.[6] However, in a recent study Wintoki et al. (2009) have shown that there is another source of endogeneity which they termed as dynamic endogeneity. Dynamic endogeneity refers to a situation that arises because of the fact that the relations among a firm's observable characteristics are likely to be dynamic. For example, in the insider ownership–performance relationship, current ownership levels will affect future performance and these may, in turn, affect future ownership levels. Thus, without recognizing dynamic endogeneity the insider ownership–performance relationship would produce biased results. Wintoki et al. (2009) show that a dynamic panel GMM estimator not only eliminates any bias that may arise from ignoring dynamic endogeneity, but also provides powerful instruments that account for simultaneity and unobservable heterogeneity. In particular, the system GMM estimator developed by Arellano and Bover (1995) and Blundell and Bond (1998) is shown to produce efficient estimates while controlling for unobserved heterogeneity, simultaneity and dynamic endogeneity (Wintoki et al. 2009). In this subsection we apply system GMM estimation of dynamic panel by using 'xtabond2' command in STATA program (Roodman 2009), as a robustness check for our earlier results.

The specification we use for system GMM estimation takes the following general form:

$$\text{Perf}_{it} = \alpha_0 + \alpha_1 \text{Perf}_{it-1} + \alpha_2 \text{OWNINS}_{it} + \alpha_3 \text{Control variables} + \eta_i + \varepsilon_{it} \ldots \quad (7)$$

where η_i is the unobserved firm-specific effects and ε_{it} is the error term. Unobservable characteristics of the firm that have a significant effect on firm performance are captured in η_i.

[5]This section draws from the author's previously published work titled 'Does capital structure depend on group affiliation? An analysis of Indian firms' published in Journal of Policy Modeling, 2013.

[6]Unobservable heterogeneity arises if there are unobservable factors that affect both the dependent and explanatory variables. On the other hand, simultaneity arises if the independent variables are a function of the dependent variable.

We report the system GMM estimation results for four alternative specifications in Table 5. We carry out two-step GMM estimation, since it is more efficient than one-step estimation. Also, the Sargan over-identifying restriction is heteroscedasticity-consistent only if it is based on the two-step estimation (Arellano and Bond 1991; Blundell and Bond 1998). The efficiency of the GMM estimator, however, depends on the assumption that the dependent and other explanatory variables are valid instruments and the error terms do not exhibit serial correlation. To address these issues, Arellano and Bond (1991) proposed three tests. The first is to test the hypothesis that there is no first-order serial correlation of the error term. Under the null hypothesis of no serial correlation, the test statistic is distributed as a standard normal. The second is to test that there is no second-order serial autocorrelation of the error term, which is distributed as a standard normal under the null hypothesis of no serial correlation. The third is the Sargan test of over-identifying restrictions. This tests the validity of the instruments and is asymptotically distributed as χ^2 under the null of instrument validity.

We observe from Table 5 that the Sargan test reveals the acceptance of the null hypothesis of instruments validity for all the four alternative model specifications. This indicates that it is appropriate to treat firm-specific characteristics as exogenous. The test statistic for first-order serial correlation, applied to the differenced residuals, shows that it is significant in all the models, which is expected (Mileva 2007). On the other hand, the second-order serial correlation, which is more important because it detects autocorrelation in levels, is not significant in models (1) and (4), indicating the models are not misspecified. However, it is significant in models (2) and (3) and hence raises doubts about their correct specifications.

We now interpret the estimates of the coefficients for the alternative models. In model (1), the variable OWNINS, our central independent variable of interest, is negative and insignificant. In models (2) and (3) we added OWNINS^2 and OWNINS^3, respectively. In both the cases, the coefficients on these variables are statistically insignificant. Then in model (4) we include the term OWNINS^4 and observe that OWNINS is positively significant at 5% level, OWNINS^2 is negatively significant at 5% level, OWNINS^3 is positively significant at 10% level, and OWNINS^4 is negatively significant at 10% level. Model (4) appears to be the best model based on the Wald Chi-square statistics. In the dynamic panel model based on GMM estimator, Wald Chi-square statistics should be used to decide on the selection of the optimal model (Candelon et al. 2012). Therefore, the system GMM estimation also supports a quartic relationship between insider ownership and firm performance. It suggests that our semi-parametric panel regression results are robust.

As model (4) is the best model, we interpret the results for the other control variables. We find that SIZE is negative and significant, GROWTH is positive and significant, LEV is negative and significant, R&D is positive but insignificant, and AGE is negative and significant in model (4). Therefore, these results are also in conformity with the semi-parametric panel regression results.

Table 5 Estimation results of system dynamic panel regression for Tobin's q for Indian firms

	Model (1)	Model (2)	Model (3)	Model (4)
Constant	1100465 (697149.2)	911417.4 (850629.1)	151058.7 (1259963)	−1855064 (1698239)
Tobin q_{it-1}	0.146 (0.122)	0.098 (0.138)	0.074 (0.132)	0.0658992 (0.131)
SIZE	49659.92 (36898.49)	48167.53 (40117.18)	54028.78 (40524.86)	−38260.58 (16461.76)*
GROWTH	765.934 (574.237)	786.898 (625.798)	864.201 (605.641)	1402.401 (680.289)**
LEV	−904486.4 (303721)**	−1017071 (317161.5)*	−1061947 (299070.8)*	−1089297 (308396.3)*
R&D	101.095 (90.207)	121.706 (106.04)	134.067 (111.038)	118.497 (91.195)
AGE	−153762.5 (125503.5)	−212910.5 (129913.1)***	−238623.2 (133917.9)***	−189232.98 (62961.03)*
OWNINS	−3167.117 (6241.7)	24772.82 (32263.03)	90574.02 (76894.64)	436503.7 (224342.2)**
$OWNINS^2$		−328.476 (340.833)	−1751.302 (1511.053)	−16842.73 (8877.07)**
$OWNINS^3$			8.992 (9.524)	254.561 (141.317)***
$OWNINS^4$				−1.347 (0.786)***
AR(1)	−2.28**	−1.66***	−2.19**	−1.89***
AR(2)	−1.45	−2.34**	−2.27**	−1.52
Sargan test (df)	947.92 (12)	875.95 (13)	853.19 (14)	743.21 (15)
Wald χ^2 (df)	35.38 (7)	37.08 (8)	39.40 (9)	41.38 (10)

Note The same as Table 3

7 Conclusion

This paper examines the relationship between insider ownership and firm performance in India for the period 2006–2013 and provides some new insights. Analysing the characteristics of corporate firms in India, predominated by business groups, our theoretical predictions hypothesize a quartic relationship between insider ownership and firm performance. Our empirical findings also confirm such a nonlinear relationship between insider ownership and firm performance. We have first applied semi-parametric panel fixed effect regression method and then applied piecewise panel regression method following Morck et al. (1988). As the semi-parametric panel regression method does not impose any prior functional specification, the true nature of the relationship between insider ownership and firm performance could be visualized from the data itself. The inflexion points in the piecewise regression are

identified by looking at the graphical plots obtained from the semi-parametric panel regression method which could be methodologically better than Morck et al. (1988).

Our results for Tobin's q show a quartic relation with insider ownership based on semi-parametric panel regression method. This is in contrast to earlier studies which have reported a cubic or quadratic specification and confined it to one maximum, by construction. The empirical relationship between insider ownership and firm value implies that ownership matters and that a change in share ownership by insiders can be used to change firm value. The relationship between insider ownership and firm value implies that, if all firms are operating at optimum ownership, then in a cross section of firms, 'firm value is a function of the structure of equity ownership' (McConnell and Servaes 1990). McConnell et al. (2007), however, argue that all firms do not operate at this optimal ownership structures all the time because when insiders trade, stock prices increase on average, but those negative stock price changes are concentrated over a specific range of share ownership. In other words, negative stock price reactions occur when initial inside ownership is very high.

McConnell et al. (2007) cite studies which evidence that the company board informally coerced top-level managers into buying more stock, and the firms' accounting and stock price performance improved. Thus, they argue that the purchase by insiders that we observe may simply be trades pushing insider ownership towards its optimal level.

Our observed quartic relationship between insider ownership and Tobin's q is the result of two opposing forces, viz. the alignment effects of ownership and the entrenchment effects that arise from a desire to maintain control. In our study we get two maxima: one at the 20–25% level and the other at the 50–80% level. The first maximum occurs when family members are involved in the management of the company and are interested in maximizing firm value. At the second maximum, the company board will force the managers to buy shares of the company to increase firm value, as discussed in McConnell et al. (2007).

To test the robustness of our results we have also applied the dynamic panel regression method which takes care of the problem of endogeneity. The dynamic panel regression results are in conformity with the semi-parametric regression results, leading to a quartic relationship between insider ownership and firm performance. While taking care of the problem of endogeneity, the dynamic panel regression results show that insider ownership is a determinant of firm performance and the relationship is nonlinear and hence these findings in the Indian context are in contrast to the findings by Himmelberg et al. (1999) who have found no relationship between insider ownership and performance. However, the application of piecewise panel regression method gives an insignificant result in our study.

Our research shows that the relationship between insider ownership and firm performance depends on the particular characteristics of corporate firms in an economy. Thus, the earlier results established for the developed countries like the USA and the UK may not always be extended to other countries like India.

References

Anderson, R. C., & Reeb, D. M. (2003). Founding-family ownership and firm performance: Evidence from the S&P 500. *Journal of Finance, 58*(3), 1301–1328.

Arellano, M., & Bond, S. (1991). Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *Review of Economic Studies, 58*(194), 277–297.

Arellano, M., & Bover, O. (1995). Another look at the instrumental variable estimation of error-components models. *Journal of Econometrics*, *68*(1), 29–51.

Baltagi, B. H., & Li, D. (2002). Series estimation of partially linear panel data models with fixed effect. *Annl Econ and Fin, 3,* 103–116.

Berle, A. A., & Means, G. C. (1932). *The modern corporation and private property*. New York: MacMilan.

Bertrand, M., Mehta, P., & Mullainathan, S. (2002). Ferreting out tunnelling: An application to Indian business groups. *Quarterly Journal of Economics, 117*(1), 121–147.

Bloom, N., & Van Reenen, J. (2007). Measuring and explaining management practices across firms and countries. *Quarterly Journal of Economics, 122*(4), 1351–1408.

Blundell, R., & Bond, S. (1998). Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics, 87*(1), 115–144.

Candelon, B., Ahmed, J., & Straetmans, S. (2012). Predicting and capitalizing on stock market Bears in the U.S. Working Paper No. RM/12/019, Maastricht University.

Chakrabarti, R., Megginson, W. L., & Yadav, P. K. (2007). Corporate governance in India. CFR Working paper no. 08-02. http://hdl.handle.net/10419/41393.

Chibber, P. K., & Majumdar, S. K. (1999). Foreign ownership and profitability: Property rights, control and the performance of firms in Indian industry. *Journal of Law and Economics, 42*(1) (Part 1), 209–238.

Cho, M. H. (1998). Ownership structure, investment and the corporate value: An empirical analysis. *Journal of Financial Economics, 47,* 103–121.

Chua, J. H., Crisman, J. J., & Sharma, P. (2003). Succession and nonsuccession concerns of family firms and agency relationship with nonfamily managers. *Family Business Review, 16*, 89–107.

Cui, H., & Mak, Y. T. (2002). The relationship between managerial ownership and firm performance in high R&D firms. *Journal of Corporate Finance, 8*, 313–336.

Davies, J. R., Hillier, D., & McColgan, P. (2005). Ownership structure, managerial behaviour and corporate value. *Journal of Corporate Finance, 11*, 645–660.

DeAngelo, H., & DeAngelo, L. (2000). Controlling shareholders and the disciplinary role of corporate payout policy: A study of the Times Mirror company. *Journal of Financial Economics, 56*(2), 153–207.

Demsetz, H., & Lehn, K. (1985). The structure of corporate ownership: Causes and consequences. *Journal of Political Economy, 93*, 1155–1177.

Demsetz, H. (1983). The structure of ownership and the theory of the firm. *Journal of Law and Economics, 26*, 375–390.

Denis, D., & Denis, D. (1994). Majority owner-managers and organizational efficiency. *Journal of Corporate Finance*, *1*(1), 91–118.

Facccio, M., & Lasfer, M. A. (1999). Managerial ownership, board structure and firm value: The UK evidence. Working paper.

Fama, E. F., & Jensen, M. C. (1983). Separation of ownership and control. *Journal of Law and Economics, 26*, 301–325.

Fama, E. F. (1980). Agency problems and the theory of the firm. *Journal of Political Economy, 88*, 288–307.

Gopalan, R., Nanda, V., & Seru, A. (2007). Affiliated firms and financial support: Evidence from Indian business groups. *Journal of Financial Economics, 86*, 759–795.

Hermalin, B., & Weisbach, M. S. (1991). The effect of board composition and direct incentives on corporate performance. *Financial Management*, *20*, 101–112.

Himmelberg, C. P., Hubbard, R. G., & Palia, D. (1999). Understanding the determinants of managerial ownership and the link between ownership and performance. *Journal of Financial Economics, 53,* 353–384.

Holderness, C. G., & Sheehan, D. P. (1988). The role of majority shareholders in publicly held corporations: an exploratory analysis. *Journal of Financial Economics, 20*(1–2), 317–346.

Hung, J., & Chen, H. (2009). Minimum shareholding requirements for insiders: Evidence from Taiwanese SMEs. *Corporate Governance: An International Review, 17*(1), 35–46.

James, H. S. (1999). Owners as manager, extended horizons and the family firm. *International Journal of the Economics of Business, 6*(1), 41–55.

Jensen, M. C., & Meckling, W. (1976). Theory of the firm: Managerial behaviour, agency costs and capital structure. *Journal of Financial Economics*, *3*, 305–360.

Jensen, M. C., & Murphy, K. J. (1990). Performance pay and top-management turnover. *The Journal of Finance*, 3–21.

Jensen, M. C. (2000). Theory of the firm: Governance, residual claims and organizational forms. Cambridge: Harvard University Press.

Khanna, T., & Palepu, K. (1999). Emerging market business groups, foreign investors and corporate governance. National Bureau of Economic Research Working Paper no. W6955, Cambridge, MA.

Kole, S. (1996). Managerial ownership and firm performance: Incentives or rewards? *Advances in Financial Economics, 2,* 119–149.

Lemmon, M. L., & Lins, K. V. (2003). Ownership structure, corporate governance and firm value: Evidence from the East Asian Financial Crisis. *The Journal of Finance, 58*(4), 1445–1468.

Loderer, C., & Martin, K. (1997). Executive stock ownership and performance: Tracking faint traces. *Journal of Financial Economics*, *45*, 223–255.

Mathew, S. J. (2007). Hostile takeovers in India: New prospects, challenges and regulatory opportunities. *Columbia Business Law Review, 2007*(3), 800–843.

McConnell, J. J., & Servaes, H. (1990). Additional evidence on equity ownership and corporate value. *Journal of Financial Economics,* 27: 595–612.

McConnell, J. J., Servaes, H., & Lins, K. V. (2007). Changes in market values and analysts' EPS forecasts around insider ownership changes. www.faculty.london.edu/hservaes/changes.pdf.

Mehran, H. (1995). Executive compensation structure, ownership, and firm performance. *Journal of Financial Economics, 38,* 163–184.

Mileva, E. (2007). Using Arellano-Bond dynamic panel GMM estimators in Stata. www.Fordham.edu/Elitz_usingArellano-bondGMMestimators.pdf/.

Morck, R., & Yeung, B. (2004). Family control and the rent-seeking society. *Entrepreneurship: Theory and Practice, 28*(4): 391–409.

Morck, R., Stangeland, D. A., & Yeung, B. (2000). Inherited wealth, corporate control and economic growth: the Canadian disease. In R. Morck (Ed.), *Concentrated corporate ownership* (pp. 319–369). Chicago: University of Chicago Press.

Morck, R., Wolfenzon, D., & Yeung, B. (2005). Corporate governance, economic entrenchment and growth. *Journal of Economic Literature, 43,* 657–722.

Morck, R., Shleifer, A., & Vishny, R. W. (1988). Management ownership and market valuation: An empirical analysis. *Journal of Financial Economics, 20*, 293–315.

Mudambi, R., & Nicosia, C. (1998). Ownership structure and firm performance: Evidence from the UK financial services industry. *Applied Financial Economics, 8*, 175–180.

Newson, R. (2001). B-splines and splines parameterized by their value at their reference points on the x-axis. *Stata Technical Bulletin, StataCorp LP, 10*(57).

Pattanayak, M. (2001). Insider ownership and firm value: Evidence from Indian corporate sector. Working paper.

Richter, A., & Chakraborty, I. (2015) Promoter ownership and performance in publicly listed firms in India: Does group affiliation matters? Occasional Paper No. 45, Institute of Development Studies Kolkata, Kolkata.

Roodman, D. (2009). How to do xtabond2: An introduction to difference and system GMM in Stata. *Stata Journal, 9*(1), 86–136.

Sarkar, J., & Sarkar, S. (2000). Large shareholder activism in corporate governance in developing countries: Evidence from India. *International Review of Finance, 1*(3), 161–194.
Selarka, E. (2005). Ownership concentration and firm value. A study from Indian corporate sector. *Emerging Markets Finance and Trade, 41*(6), 83–108.
Shleifer, A., & Vishny, R. W. (1997). A survey of corporate governance. *Journal of Finance, 5*(4), 341–368.
Short, H., & Keasey, K. (1999). Managerial ownership and the performance of firm: Evidence from UK. *Journal of Corporate Finance, 5,* 79–101.
Volpin, P. (2002). Governance with poor investor protection: Evidence from top executive turnover in Italy. *Journal of Financial Economics, 64*, 61–90.
Wang, D. (2006). Founding family ownership and earnings management. *Journal of Accounting and Research, 44*, 619–656.
Wintoki, J., Linck, J., & Netter, J. (2009). *Endogeneity and the dynamics of corporate governance*. Paper presented in European Summer Symposium in Financial Markets, July 13–24, 2009, Study Centre Gerzensee.
Wruck, K. (1988). Equity ownership concentration and firm value: Evidence from private equity financings. *Journal of Financial Economics, 23,* 3–28.

Part IV
Innovation and Management

Generic framework of a Business Incubator Model for a Sustainable Innovation Ecosystem

R. A. Bhaskar and B. V. Phani

1 Introduction

Business enterprises/industries[1] are the engines which propel NEG. NEG in terms of industry efficiency depends upon the optimum utilization of available resources both domestic and non-domestic. Economies in various states of progress strive to optimize the utilization of these resources by increasing the efficiency of BEs. This is mostly achieved through a Structured & Directed Approach (SDA) of intervention in Science, Technology & Innovation (STI) (Fig. 1). These interventions through SDA are dogged with high probability of market, systemic and translational failure. The literature identifies three primary causes for market failures, namely, **e**xternalities, uncertainty and indivisibilities leading to lowering the investment in R&D, knowledge and technological innovation (UNCTAD 2011). This in turn results in lowering the productivity of BEs in both short and long terms (Crespi 2012). Systemic failure refers to the inability of the firms to leverage the interrelationship of various stakeholders in the innovation ecosystem and the institutional design. Translational failure refers to the inability of these interventions to effectively transition, socialize and monetize the knowledge created. These failures or weaknesses would lead to dissonance in the interaction of innovation actors (universities, BEs, government agencies, and research institutes) impeding the leverage and assimilation potential of the available knowledge pool for the greater benefit and sustained economic growth (UNCTAD 2011).

[1] Business enterprises/industries include all type of services, manufacturing, knowledge, tangible and intangible resource processing centers/value addition centers etc.

R. A. Bhaskar · B. V. Phani (✉)
Indian Institute of Technology (IIT) Kanpur, Kanpur, India
e-mail: bvphani@iitk.ac.in

R. A. Bhaskar
e-mail: rbhaskar@iitk.ac.in

P. Ray et al. (eds.), *Economics, Management and Sustainability*,
https://doi.org/10.1007/978-981-13-1894-8_12

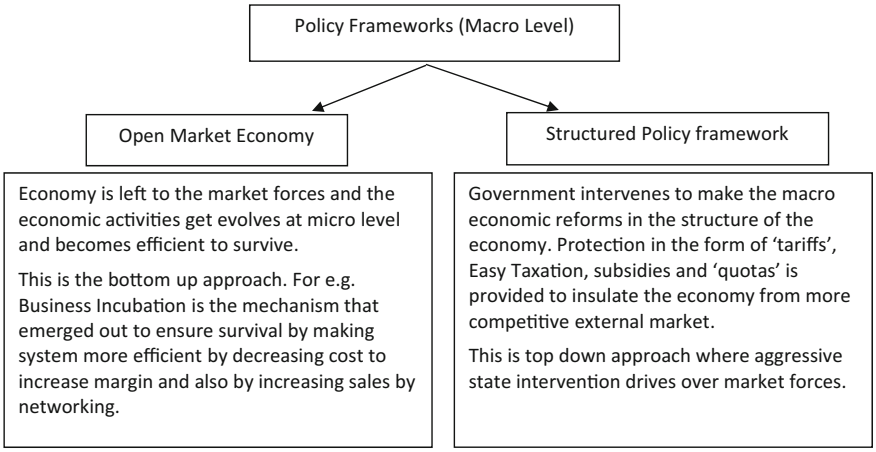


Fig. 1 Policy frameworks. *Source* Authors' own

Implicit policy interventions are intended to ensure the survival and efficient functioning of these enterprises by nurturing and safeguarding them from an exploitative and dynamic external environment. Implicit policy instruments include subsidies, quota, tariffs, trade policy, taxation which address the problem of market failure. But both direct and indirect financial support policies, i.e., subsidies and fiscal incentives, just replace business-funded innovation activity resulting in crowding out effect (Montmartin and Herrera 2015) and are not addressing the fundamental issues leading to the same. Explicit and active policy tools focus on ameliorating innovation ecosystem by developing STI infrastructure, through increased public spending on R&D, promoting entrepreneurship and endeavoring to make the domestic enterprises competitive to address their market failure probability and also mitigate the interaction interface failure between the ecosystem stakeholders.

This resulted in the emergence of a novel mechanism of a state-supported formal business incubation model leveraging both explicit and implicit policy intervention accentuating the efficiency of the innovation ecosystem addressing both systemic and market failure. This mechanism is one of the components of a quintuple helix interface model encompassing and leveraging the strength of multiple actors within the innovation ecosystem for efficient knowledge assimilation and translation.

This novel mechanism is intended to mitigate the issues faced by New Business Enterprises[2] (NBE) due to unproven technologies and business models in their formative years. These issues as identified from the literature could vary from lack of domestic demand, technology disruption and failure, access to finance and financial expertise, lack of capital, acquisition and managing of resources, lack of managerial competencies, etc. The literature concludes that BIs provide a minimum of four

[2]For the purpose of this study, NBEs refer to new enterprises with 'unproven and/or innovative' technologies and disruptive business models, whose dependence on the knowledge ecosystem is a critical factor in their survival, success and sustainability both in the short and long terms.

services: infrastructure, business support services (e.g., legal/accountancy), coaching and mentoring, funding and networking services. These services are aimed at addressing the issues mentioned above. Given this it is presumed that a BI would be the optimal solution to address and mitigate the issues faced by BEs during their formative years. This in turn is proposed to significantly reduce the mortality rate of NBEs which is currently calculated to be as high as 90%.

BIs as a formal mechanism for supporting NBEs saw its emergence in developed economies like USA and in the last two decades became gradually popular in emerging economies as a mechanism to promote innovation and entrepreneurship in line with their aspirations of becoming a globally competitive knowledge economy.

Different models of business incubation have been documented over time as implemented in different parts of the world. Similar BI models had exhibited a varied performance efficiency in different economies. Given the uniformity of service offering by BIs as indicated earlier, it is evident that one size fit all cannot be the norm. A linkage with the state of ecosystem, entrepreneurial needs and aspirations and BI needs to be identified/designed for a given socio economic system.

This study undertakes a review of BI literature to capture the BI, NBE dynamics within a BI-NBE dyad framework. This was undertaken by first freezing the definition of entrepreneur, entrepreneurial process, incubation process and then undertaking the survey of service offering provided by distinct incubator archetype (model), their selection methodology and finally factors to measure their performance to document BI best practices and proposing design and implementation course corrections for ameliorating the innovation ecosystem further.

2 Entrepreneurial Process

The first step of any business incubation process is a critical evaluation of ideas and opportunities and identifying the expectation of entrepreneurs' with respect to the BI. Since models followed by BIs span, a broad spectrum varying from offering highly generalized services and facilities, open to entrepreneurs with business ideas in any domain to highly focused approach open only to specific domains and technologies with access to limited services and facilities.

Mapping the BI models to entrepreneurial expectations necessitates a predetermined definition of an Entrepreneur. In spite of many attempts over the years to define an entrepreneur, there is no general consensus in this regard. The word 'entrepreneur' itself is derived from French 'entreprendre' meaning to begin something or undertake. The first reference to the usage of this word is attributed to Cantillon in 1700 who defined it as a person taking risk and managing a firm or business. But the word attained global prominence only in 1934 when Schumpeter defined entrepreneurship as 'creative destruction' replacing old with new ideas, technologies and models of doing business. Gartner (1989) defined entrepreneur as one who creates organization but Carland et al. (1988) differentiated entrepreneur from small business owner on the basis of scope, objective and scale of their operation.

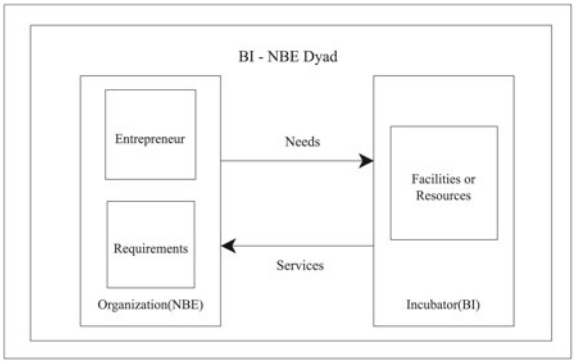


Fig. 2 BI-NBE dyad

For the purpose of this study, we stick to the definition of an entrepreneur as 'one who innovates, creates and manages an organization to generate profit and/or growth rather than personal goals (self-employment) in the face of constrained resources.'

BI-NBE dyad and its main stakeholders are shown in Fig. 2. An Entrepreneur, with certain personal characteristics attributed to any individual like creativity (to innovate), risk taking propensity, leadership skills may promote a BE to monetize an idea by gathering individuals with necessary skill-set to further this. The reverse may not be true, i.e., individuals having these characteristics need not necessarily be an entrepreneur. This formal organization in its journey toward becoming a successful BE may require support mechanisms and services at different stages of its life cycle. Existing and evolving BI models are considered to be a one stop solution to provide many of these support mechanisms and services to these BEs to improve their probability of success.

3 Business Incubation

Business incubation may be defined as "a process enacted by BIs, angels and venture capital organizations in order to facilitate the entrepreneurial process" (Hackett and Dilts 2004a). NBIA defines BI as a 'unique and highly flexible combination of business development processes, infrastructure and people designed to nurture new and small businesses by helping them to survive and grow through the difficult and vulnerable early stages of development.'

The word 'incubate' finds its origin in the Latin '*incubātus*' *or* '*incubāre*' which means to lie or recline upon, to sit upon (eggs) or to hatch. The Oxford dictionary interprets incubator as a 'an enclosed apparatus providing a controlled environment for the care and protection of premature or unusually small babies or an apparatus used to hatch eggs or grow microorganisms under controlled conditions.' Drawing a parallel with this a BI is a combination of equipment, apparatus or mechanism which insulates the 'incubate' from uncertain environmental conditions by providing a

controlled growth environment so that organization can survive and adapt itself to the environment gradually over a period of time to ensure an NBE's early stage survival.

The formal history of BI is not more than 60 years; the first incubator was opened in USA in 1959. Incubation as a viable model to support NBEs shot into limelight after the passage of Bayh-Dole Act in 1985 and formation of the NBIA[3] (now InBIA[4]). According to NBIA data, there were only 12 incubators in northern America in 1980 and the latest count shows more than 7000 active incubators/accelerators worldwide. USA and China account for more than half of the same followed by Europe. India has around 140 with more than 50% attached to academic and research institutions.

The role and criticality of NBEs supported by BIs in a nation's economic growth is well documented in the literature. Emerging economies like China, Brazil and India have also emphasized on the role of BIs in recent years. BIs in India were promoted by the National Science and Technology Entrepreneurship Development Board (NSTEDB) with the objective of improving the efficiency of translating research output for the larger benefit of the society by employing a holistic intervention policy framework. This was initiated by establishing Science and Technology Entrepreneurs Parks (STEP) in the early 1980s, and the process was then accelerated by launching of Technology Business Incubators (TBI) in early 2000.

The Indian start-up ecosystem with over 5000+ start-ups and expected to cross 10,000 mark by 2020, attracting private investments of over USD 1.8 Billion in 2017, is considered to be the 3rd largest start-up location globally (NASSCOM[5] Startup Report 2017). With a strong angel, venture capital and private equity backbone of over 500+ active investors and over 190+ incubators and accelerators, this ecosystem is poised to become a beacon for entrepreneurs with innovative ideas both within and outside the country.

Given the above and with a mortality rate of 20–25%, the criticality of BIs as a policy intervention tool for the success of this ecosystem cannot be emphasized enough. Currently there is large diversity in the configuration of the BIs to support the NBEs requirements at various stages of their lifecycle. The configuration and the services they provide depend upon incubator type, size, goal and strategic objective of the BI identified beforehand.

[3]NBIA is National Business Incubators Association

[4]InBIA is International Business Innovation Association

[5]The National Association of Software and Services Companies (NASSCOM) is a trade association of Indian Information Technology and Business Process Outsourcing industry.

3.1 Business Incubation Frameworks

The existing literature classifies business incubation in terms of affiliation (Smilor 1987), financial sponsorship (Grimaldi and Grandi 2005; Smilor 1987), geographic location (Grimaldi and Grandi 2005; Hackett and Dilts 2008; Smilor 1987; Von Zedtwitz 2003), focus and objective (Grimaldi and Grandi 2005; Hackett and Dilts 2008; Smilor 1987; Von Zedtwitz 2003), market or segmental focus (Grimaldi and Grandi 2005; Von Zedtwitz 2003) (Table 1).

Affiliation refers to Private, University, Government and Non-profit (Smilor 1987), and on the basis of Geographic Locations, they can be classified into four types, they are regional, urban, semi-urban, rural (Hackett and Dilts 2008). The types of incubators on the basis of objective are for-profit and not for-profit (Von Zedtwitz 2003) and on the basis of segmental focus are Bio-tech, Empowerment, High-tech, and Mixed use (Hackett and Dilts 2008). Grimaldi and Grandi (2005) have taken mission, sector, location, market, intervention phase and duration, revenue sources, origin of ideas, offered services and management teams as the incubator characterizing variables to segregate business incubators in the following four categories. They are 'Business Innovation Centres (BICs)'; 'University Business Incubators (UBIs)'; 'Independent Private Incubators (IPIs)'; 'Corporate Private Incubators (CPIs)'. A summary of the identified incubator archetype from various combination of characterizing variables by different authors is presented in Appendix 1: Classification.

Table 1 BI classification

Classification metrics	Incubator types	Indicative literature
Affiliation or financial sponsorship	Private University Government Community	Grimaldi and Grandi (2005), Smilor (1987), Smilor et al. (1990)
Geographical location	Urban Sub-urban Rural Regional	Grimaldi and Grandi (2005), Hackett and Dilts (2008), Smilor (1987), Von Zedtwitz (2003)
Objective	For-profit Not for-profit	Grimaldi and Grandi (2005), Hackett and Dilts (2008), Smilor (1987), Von Zedtwitz (2003)
Segmental focus	Bio-tech High-tech Mixed use Other	Hackett and Dilts (2008), Von Zedtwitz (2003)

3.2 Business Incubation Models

The literature identifies two types of incubator models as presented by different scholars (Campbell et al. 1985; Grimaldi and Grandi 2005; Hackett and Dilts 2004a; Hansen et al. 2000; Lalkaka 2001, 2002; Lalkaka and Abetti 1999; Manimala and Vijay 2012; Meyer 2003; Peters et al. 2004; Robinson 2010; Smilor 1987; Vij and Jhanji 2013; Von Zedtwitz 2003). The first focuses on the structure of incubator and the second on the process in explaining the set of services provided (Fig. 3).

The elements or the incubator characterizing variables using the structural model consists four basic elements (affiliation, objective, location and segmental focus), and for each element, there are different sub-heads. For example, affiliation (x) may be private (x_1), university (x_2), government (x_3) or community (x_4). Geographic location (y) may be Urban (y_1), Rural (y_2), Sub-urban (y_3) or Regional (y_4) (Fig. 4). Incubator objective (z) may be for for-profit (z_1) or not for-profit (z_2). Incubator may also have segmental focus (w) for biotech (w_1), high-tech (w_2), mixed use (w_3) or other (w_4). For a unique incubator, each element takes one value at a time from the available sub-heads. Each unique incubator takes only one value $x; y; z$; and w from $x_1, x_2, x_3, x_4; y_1, y_2, y_3, y_4; z_1, z_2$ and w_1, w_2, w_3, w_4, respectively. If we take the combination matrix of these elements, then there are total $\binom{4}{1} * \binom{4}{1} * \binom{2}{1} * \binom{4}{1} = 128$ combinations possible. However, all the combinations may not be viable or efficient and the ones that have better impact as identified by the literature including those categorized using a process model are presented in Appendix 2: BI Models.

The feasible or models currently in practice out of all the possible combinations are presumed to follow best practices in their given domain. In spite of this, the structural models were not performing as expected, i.e., there was significant difference in their performance even though different incubators were identified with similar structural parameters. This leads us to believe keeping everything else equal these differences

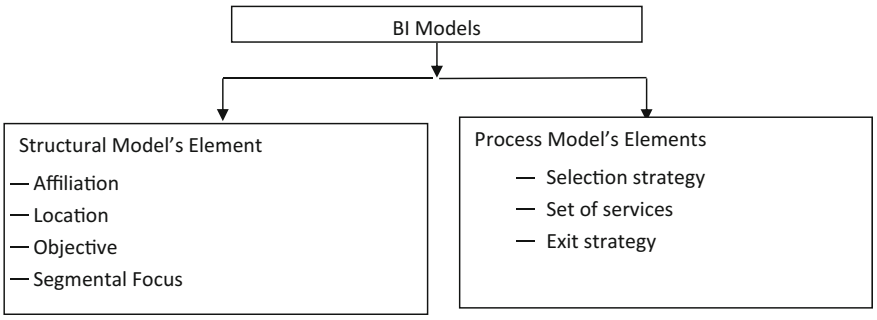


Fig. 3 Model types

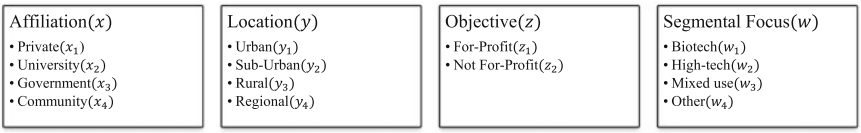


Fig. 4 Elements of incubator framework

can be accounted for by the variation in the nature and efficiency of services offered to the NBEs served by them. This brings us to seek answers in the process model categorization framework.

BI process models focus on selection strategy, set of services and exit (Bergek and Norrman 2008; Campbell et al. 1985; Hackett and Dilts 2004a; Meyer 2003; Peters et al. 2004; Smilor 1987; Von Zedtwitz 2003) (Table 2). A balanced and rigorous selection process ensures firm survival, and as a consequence it assumes critical significance for promoting specific innovation type (Aerts et al. 2007). NBEs, which already benefit from implicit policies/government intervention, are strong enough to survive by themselves as insulated, and favorable environment is already ensured to them, and BI as a mechanism will have very little room for contributing to their survival (Scillitoe and Chakrabarti 2010). The focus should be more on NBEs, which may face an adverse ecosystem and also require more lead time to become highly promising in terms of both economic and social returns (Hackett and Dilts 2004a). This will ensure optimum utilization and the rightful and equitable distribution of policy intervention.

In contradiction to the above, a review of the related literature indicates that significant emphasis on some or all of the following criteria like idea; market; product; or team; entrepreneur; etc., are used during selection. These approaches focus on strategies such as 'Picking the Winner' or 'Survival of the Fittest' (Bergek and Norrman 2008; Merrifield 1987). On the other hand, Aerts et al. (2007) identified 'market,' 'management team' and 'financial factors' as selection criteria and associated the importance of these factors with incubator performance. A strand of the literature proposed the idea of diagnosing of NBE needs and provide the necessary assistance, but not much elaboration on this approach was found, though it assumes that by doing so all potential NBEs can be turned into viable businesses (Campbell et al. 1985; Moreira and Carvalho 2015).

The literature on selection strategies mainly focuses on 'Picking the Winner' over 'Optimum and Equitable Resource Distribution'; this creates a market segment bias by preferring the firms in the profitable market segment and highly risky ideas with larger societal and economic impact remain unaddressed where innovation is the only solution, thus defeating the purpose of a BI. Even though models of BI with a specific social innovation focus are being experimented with as a vertical in the structured model framework, they are having limited success. To address this, it would be a better approach if BIs selection model primarily focuses on weak but promising ideas or NBEs and the spectrum of both tangible or intangible value addition it can

provide after mapping of the needs of NBE to the service offering available with a specific BI.

Once selected, the process model emphasizes the delivery efficiency of the service offering. An NBE's survival depends critically on the BI's ability to efficiently deliver its service offering in terms of both quality and timeliness. The literature identifies that the resources offered by a BI are mainly infrastructure and physical space (European Commission Enterprise Directorate General 2002; Grimaldi and Grandi 2005; Hackett and Dilts 2004b; Peters et al. 2004; Von Zedtwitz and Grimaldi 2006). Structural support theory asserts the importance of the basic physical infrastructure (shared office space) and technical infrastructure support (access to laboratories, workshops, work stations, high end computers and other machineries) provided to select NBEs. Studies indicate that 85% of the BI's service offerings include infrastructure support (Hansen et al. 2000).

BI's service offering to NBEs also includes coaching, mentoring and expert advice (Campbell et al. 1985; Grimaldi and Grandi 2005; Kimatu 2016; Scillitoe and Chakrabarti 2010; Treibich et al. 2013; Von Zedtwitz and Grimaldi 2006). Access to knowledge services assumes primacy in incubators affiliated with or located in academic and R&D institutions (Rothaermel and Thursby 2005a; Rubin et al. 2015; Soetanto and Jack 2016).

In addition to the above, literature indicates support services like legal, accounting, bookkeeping, secretarial services, etc., with emphasis on providing access to financial resources and network are also provided (Bruneel et al. 2012; Campbell et al. 1985; Chan and Lau 2005; Meyer 2003; Mian et al. 2016; Soetanto and Jack 2016; Wiklund and Shepherd 2005). Statistics indicate that 86% of the European incubators were providing business assistance and support services (European Commission Enterprise Directorate General 2002), and this can be safely assumed to be true even today and can be considered as a minimum global standard. NBE's requirements in terms of assistance expected from BIs have undergone significant change over the last two decades. Many services once considered to be in the specific domain of a BI are now available otherwise due to the increase in the demand side bringing forth scale economies which can be exploited.

Both 'business assistance & support services' and 'coaching, mentoring and knowledge flow' require networking and interaction strength of any BI. Social capital (McAdam and McAdam 2008) and network theory (Hansen et al. 2000) propound network strength as the most critical factor in increasing and channeling knowledge and resources flow, resulting in increased learning in terms of business knowledge and capability of NBEs (Bøllingtoft and Ulhøi 2005; Díez-Vial and Montoro-Sánchez 2016; Hansen et al. 2000; Zhang et al. 2016b). Highly networked incubators have ensured the reduction in NBE mortality rate to 5% (Hansen et al. 2000). Weak network ties furnish information and knowledge, while strong network ties provide resource assistance (Katz et al. 2004). Business assistance in terms of NBE mentoring and hand holding is offered through counseling interactions, and technical assistance in terms of NBE technical training is provided through networking (Scillitoe and Chakrabarti 2010). The 'co-production' theory emphasizes on the

intensity of business assistance in terms of time allocated by BI along with the readiness and willingness of the NBE for the success of the BI-NBE dyad (Rice 2002).

Resource-based theory suggests that the availability of tangible and intangible resources better referred as capability should be ensured to NBEs for increasing an enterprise's competitiveness resulting in development and success of the BI-NBE engagement (Choi and Shepherd 2004; Löfsten and Lindelöf 2005; McAdam and McAdam 2008; Rothaermel and Thursby 2005b; Zhang et al. 2016b).

BI's support to NBE's or NBE's exit from a BI is generally time bound as BI intervention model is designed to offer support till the NBE is capable to compete on its own with its product and service offerings becoming market ready. Minimum support time can be based on factors like the NBE's self-sustainability, fixed number of years, financial performance, i.e., revenue growth and any other factor which provides a measure of market readiness like first customer (Bruneel et al. 2012; Rothaermel and Thursby 2005a). There exists a dearth of the literature on the impact of the exit models and the performance of BI-NBE innovation system as a whole. Studies on BI performance metrics have been undertaken, and certain BI performance metrics or indicators have been identified but a consensus in this regard has not been arrived yet (Theodorakopoulos et al. 2014).

Table 4 in Appendix 3 lists Business Incubation Performance Metrics (BIPMs) as identified by various studies in this regard over time.

It has been observed that process models as presented in the literature are viewed from an isolated perspective and do not present a holistic approach (Theodorakopou-

Table 2 Elements of BI process model

Elements	Consists of	Research studies
Selection	Idea or product	Bergek and Norrman (2008)
	Team or entrepreneur	Aerts et al. (2007), Bergek and Norrman (2008), Merrifield (1987)
	Market orientation	Aerts et al. (2007)
Services	Physical space	European Commission Enterprise Directorate General (2002), Grimaldi and Grandi (2005), Hackett and Dilts (2004b), Peters et al. (2004), Von Zedtwitz and Grimaldi (2006)
	Coaching and mentoring	Advice Campbell et al. (1985), Grimaldi and Grandi (2005), Kimatu (2016), Scillitoe and Chakrabarti (2010), Treibich et al. (2013), Von Zedtwitz and Grimaldi (2006)
	Support services	Chan and Lau (2005), Bruneel et al. (2012), Soetanto and Jack (2016), Mian et al. (2016)
	Networking	Bøllingtoft and Ulhøi (2005), Díez-Vial and Montoro-Sánchez (2016), Hansen et al. (2000), Soetanto and Jack (2016), Zhang et al. (2016a), Meyer (2003b)
Exit	Firm age	Bruneel et al. (2012), Rothaermel and Thursby (2005a)
	Revenue	

los et al. 2014) regarding the optimal strategies (selection or exit) and diversity of service offerings. Most of the studies do not have rigorous theoretical underpinning (Hackett and Dilts 2004b). However, the existing literature in this domain indicates a general consensus regarding the structural model framework but the same is not true with respect to the process model frameworks. The current view regarding the effectiveness of BI is veering more and more toward the criticality of the 'context and implementation' effectiveness (Chan and Lau 2005).

4 Discussion and Conclusion

The criticality of the BI-NBE dyad for a nation's economic growth and sustainability cannot be emphasized enough for a knowledge economy. A review of the literature pertaining to the innovation landscape and related ecosystem brings forth the importance of a holistic approach while designing policy intervention frameworks for accelerating this ecosystem. Current approaches of treating BIs and NBEs independently have a structural weakness, and it is proposed that an approach of treating BI-NBE as a dyad would result in better efficiencies. To support the same a comprehensive review of whole business incubation process and its placement as a policy tool in innovation ecosystem has been undertaken. A detailed review of the structural and process frameworks including selection, services and exit also indicates the BI-NBE dyad approach toward policy design would be appropriate. This approach has the additional advantage of integrating the personal and professional character of an entrepreneur (intent, content and character) along with the innovation strength brought by these individuals to this ecosystem through NBEs.

Our findings suggest that a holistic attempt was not undertaken to ascertain what a business incubation process should include or how it should work. However, there is a common consensus on the criticality of the availability of knowledge and finance resources. The mechanism and efficiency of delivery of these critical service offerings which keep changing over time is considered to be a dynamic process and an optimal approach or model is yet to be arrived at. The role of network strength in value addition within the BI-NBE dyad has been identified as crucial for the success of this ecosystem in terms of growth and impact. It is proposed that network strength as defined by the BI's ability to provide knowledge and business services has a very narrow interpretation, it is proposed that policy interventions in terms of fostering BI-BI network linkages in leveraging the strengths of individual BIs to the benefit of other BIs as and when required would have an accelerating effect on the growth and impact of this ecosystem. The current approach of leveraging network strengths in terms of vertical integration of BI, Universities & Accelerators based on the life stage of the NBE and horizontal in terms of individual BI service offerings should be augmented by leveraging BI-BI network strengths through a suitably designed policy intervention to reduce the 'reinventing the wheel' approach currently being followed.

Appendix 1: Classification of Incubators

Business incubation can be classified based on affiliation and financial sponsorship, geographic location and focus or objective (Smilor 1987).

Grimaldi and Grandi (2005) identified mission, sector, location, market, intervention phase and duration, revenue sources, origin of ideas, offered services and management teams as characterizing variables. They classified BI on the basis of affiliation or financial sponsorship into four types, they are:

Business Innovation Centers (BICs): These are research and development agencies sponsored by governments and non-profit groups responsible for allocating funds. They are having collaboration with various organizations such as academic and research institutes. Their primary objective is to create jobs for revitalizing economically backward areas.
University Business Incubators (UBIs): They are affiliated to or sponsored by the universities for accessing their research laboratories and other facilities. They are usually non-profit incubators whose primary objective is to commercialize university research by promoting or forming economically viable ventures. University provides access to their research laboratories, equipments, computers, libraries, assistance and the expertise of faculties.
Independent Private Incubators (IPIs): These are for-profit private incubators and are not receiving any financials aids or grants from government or any state agency. They are economically self-sustainable. They provide shared facilities like space, equipment and other business services.
Corporate Private Incubators (CPIs): These incubators are backed by some corporate business houses. They recruit new ventures to sell them services and invest into them. Not only they provide them shared business facility but also buy other services required for new incubatee.

Hackett and Dilts (2008) classified BI on the basis of geographic locations into four types: they are regional, urban, sub-urban, and rural. The types of incubators on the basis of focus or objective are traditional incubator, technology, cultural, social, agricultural, mixed.

According to Von Zedtwitz (2003), BI services and the way they are managed are affected by incubator archetype. Five incubator archetypes identified on the basis of the scope and objective are shown in Table 3: Incubator Archetype.

Appendix 2: BI Models

A thorough review of the literature has been done to find the critical services provided by incubators. Various models of business incubation studied were found to have different components, some of which were overlapping. The model components were basically the services provided by BIs.

Table 3 Incubator archetype

Strategic	Incubator archetype	Scope
For-profit	University	Segment focus
	Regional business	Geography, segment focus, industry
Not for-profit	Company internal	Geography, segment focus
	Independent commercial	Geography, segment focus, industry
	Virtual	Segment focus

Campbell et al. (1985) proposed a model of business process incubation in the early years when incubator–incubation concept was taking its shape (Fig. 5). They attempted to conceptualize whole incubation process into four stages of incubation which adds value to potential incubate:

Stage 1: A group of business expert diagnoses the business needs of new incubatee from their collective experience.
Stage 2: Selection of the new incubatee is done for those who can be provided with cost effective services. The monitoring, coordination and implementation of the services needed by the incubatee are provided cost effectively to increase the competitiveness and reduce the chances of failure.
Stage 3: Incubator coordinates in fulfilling the capital need for product development and other third party services.
Stage 4: They provide access to their growing network of experts and other business clients to ensure the channelizing of knowledge and information.

This model does not have clear selection criteria, and it assumes that all potential business can be turned into viable firm because of the consideration of only internal sources of value addition in the model (Moreira et al. 2012).

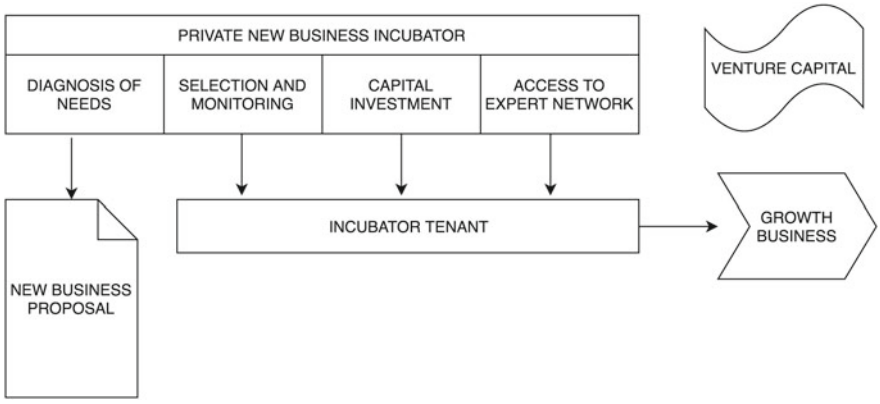


Fig. 5 Business incubation model by Campbell et al. (1985)

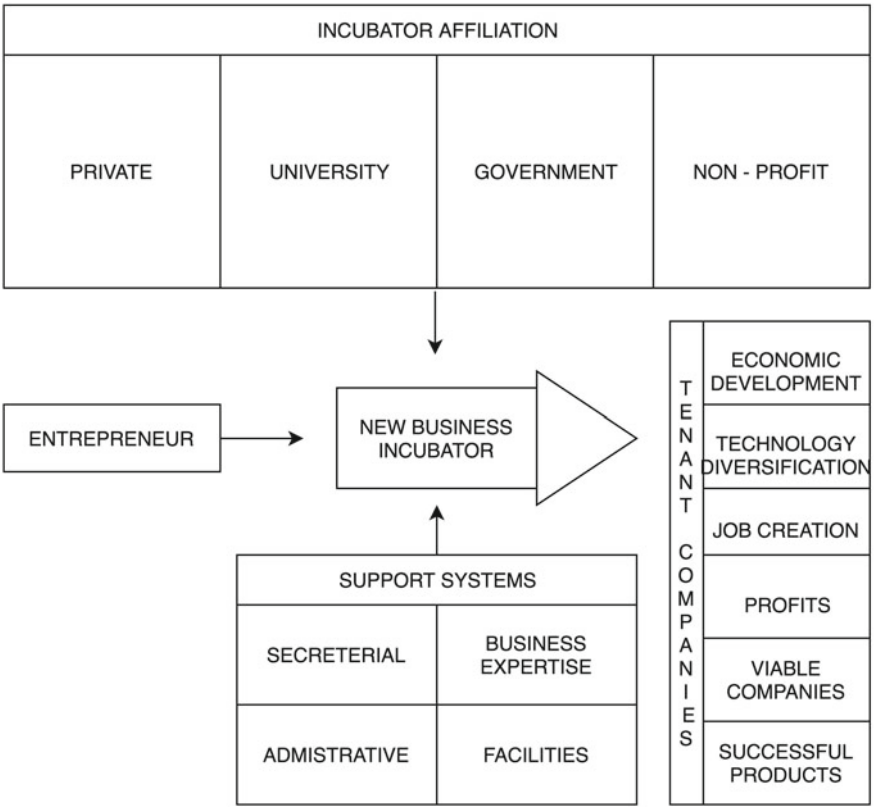


Fig. 6 Incubation model by Smilor (1987)

Smilor (1987) has gone one step further and provided organizational structure. He identified factors for success of incubators by classifying incubators according to their affiliation; the proposed model is shown in (Fig. 6). He had taken external sources of value addition, while interval was neglected.

In the proposed model, Smilor considered credibility development, entrepreneurial learning curve shortening, quick problem solving and entrepreneurial network access as the external sources of value addition. He identified following factors as critical for success:

- Tenants selection process and success perception
- Access to network of entrepreneurs and business expertise on firm site
- In-kind financial support and access provided to capital and financing.
- Coaching and education of entrepreneurs
- Incubator affiliation and Community support
- Concise program milestones with clear policies and procedures.

BI model proposed by Bergek and Norrman (2008) consists of three stages, viz. selection, business support and mediation. He has considered both sources of value addition, i.e., internal and external (Fig. 7).

The identified components to translate the process of incubation into external and internal variable are selection, business support and mediation.

Selection: According to Bergek and Norrman (2008), it is one of the important components which earlier studies neglected. Different selection criteria, such as whom to select or reject, identified under his study are:

- Primarily focused on idea: technical and business feasibility study of the idea is carried out, so only the business ideas that are having potentials got selected. And incubator ensures that no efforts go in vain.
- Primarily focused on the entrepreneur or the team: Incubator should have experts or ability to judge entrepreneur's behavioral and personal traits and ability related to business.
- 'Picking the Winner': Incubator must have competences to judge both idea and team to select potential successful venture.
- 'Survival of the Fittest' approach: In this approach, incubator does not have any selection related expertise and they select all applicants and the fittest will survive; others shut down their operation. According to the options theory of business incubation (Hackett and Dilts 2004b), if incubatee shuts their operation too early before investing too much money and effort, it is also success.

Business support includes training of entrepreneurs, advice related to business development, support related to general business matters such as book keeping, judicial matters, marketing and assistance related to finance. The strategies followed by incubators in supporting may differ according to the level of intervention they are opting for. Bergek and Norrman (2008) identified reactive and episodic counseling, proactive and episodic counseling, i.e., strong intervention and continual and proactive counseling as strategies of business support.

Mediation: Business incubator must have a network of entrepreneurs and other agents for mediating or to ensure the free flow of knowledge and technology between incubatees and innovation systems. Mediation also ensures flowing of finance and human capital and other resources to incubatees.

Hackett and Dilts's model (2004b) also considers both internal and external components and is more accepted model of business process incubation. It is result oriented and measures the process in terms of business incubator performance (BIP). The model proposed is backed by real options theory (Fig. 8).

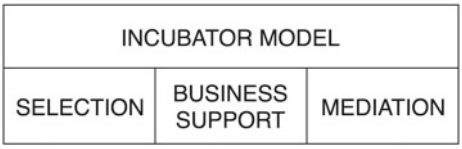


Fig. 7 Incubation model by Bergek and Norrman (2008)

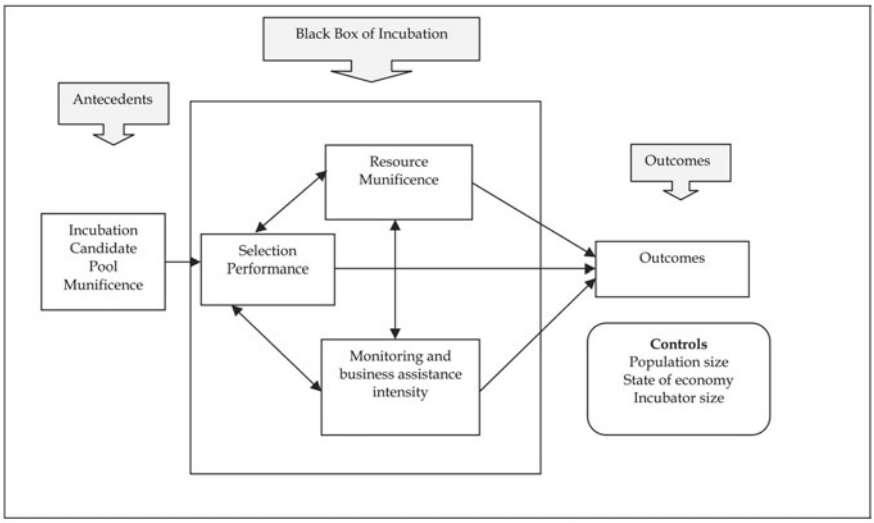


Fig. 8 Hacket and Dilt's model (2004b) model of BI

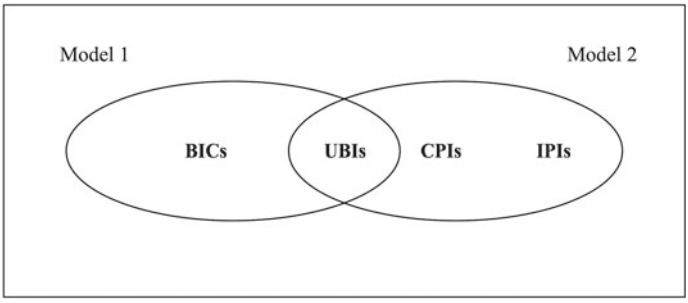


Fig. 9 Two models of BI by Grimaldi and Grandi's (2005)

The model suggests that incubator creates value by doing right selection, monitoring, assisting in business and ensuring the availability of resources. By doing selection, incubator creates an option and seeks opportunity to invest in weak firm having high potential. The subsequent investment comes in the form of resources supplied and cost associated with monitoring and business assistance. The model judges its effectiveness by measuring the end result of incubation.

Grimaldi and Grandi (2005) mapped four categories of business incubators to the two models of business incubation (Fig. 9).

BIs with Model 1 provides physical assets, they ensures accessibility to funding, expertise and other competencies that are not available in-house. They have visibility for a mid-long term orientation. Their services are oriented toward tangible assets with a mid-long term duration.

Model 2 has services more intangible and with a short-term time orientation. They generally provide access to knowledge and capital available inside as well as with external partners.

Bøllingtoft and Ulhøi (2005) seconded the concept of business incubation **c**alled networked incubator based on the social capital theory. The networked incubator is bottom-up approach, in which incubators are managed by new incubatees.

Peters et al. (2004) proposed a model to identify the role of incubators. The model has identified the services infrastructure, coaching and network which impacts the graduation rates of tenant companies. It was further improved by data analysis and interviews with incubator managers that the coaching and network services only are impacting the incubator performance. On the similar lines, Meyer (2003) suggested that more supportive and incubators with network facility were more successful.

Von Zedtwitz (2003) proposed a generic model of business incubation. He suggested that most of the incubators provide these five services. The first in line is 'access to physical infrastructure' which provides office space and IT resources. The second is office support services under the aegis of which services related to computer network, bookkeeping, security systems, IT trouble-shooting and mail services. Access to capital is one of the critical services provided by incubator. This includes angel investment, venture capital and seed funding. Incubators also provide business process support services like consulting, legal advice and book keeping. They also provide mentoring and coaching to NBE. The most critical service of all is networking services. New incubatees usually don't have network of internal as well as external clients. So incubator mediates in between and ensures that the new venture gets the resources needed.

Robinson (2010) proposed evolutionary development of business incubation model, which has three stages:

(1) Pioneering Stage: BIs directly copy the global best practices followed and replicate it to get funding from government and NGOs working for poverty alleviation and economic development. Initial entry of the incubatees and developing their business concept are the prime concern of these BIs.
(2) Operating Stage: They select high ambitious entrepreneurs to develop their product for commercialization. They obtain funds for incubatees to foster growth.
(3) Advanced Stage: They are the big size BIs incubating more firms to provide them intellectual property protection and initial capital investment. They are having strong venture team of specialized staff providing sophisticated services and are generally specialized in particular industry.

Manimala and Vijay (2012) provided framework for technology business incubators (TBIs). He defined TBI as a specific type of incubator which provides a range of services including physical infrastructure, management support, technical support, access to financing, legal assistance and networking. Vij and Jhanji (2013) provided an extensive literature review to identified three types of revenue models; they are

'landlord model,' equity position and sponsor funding. Incubators use a combination of these three. Their main function is to provide new incubatees with capital, infrastructure and network.

Appendix 3: Business Incubation Performance Metrics (BIPMs)

Key performance indicators are the mirror which depicts the true picture of the efficiency of BI models. Therefore, to find out the effectiveness of incubation models identified earlier and to solve our research question, i.e., critical factors (services) required for incubator's better performance, we first need to understand BIPIs.

According to Hackett and Dilts (2004b), 'Job growth by incubate,' 'Financial performance of incubatee' and 'Developmental advances of Incubatee at incubator exit time' are the measures of the incubator's performance.

The ratio of the 'Number of incubatees graduating from incubator' to the 'Number of incubatees shutting down their operations while being a tenant' can also be used as a metric of incubator success (Allen and Weinberg 1988).

Increases in number of jobs or growth in sales over time have been identified as growth measures, while 'product innovation,' 'the number of new patents registered' and 'strategic alliances formed' over time are the developmental measures (Udell 1990).

Hamdani (2006) mentioned the following metrics as indicators of BI success:

- The number of the jobs or the growth in employment NBEs created.
- NBE's performing R&D and the number of new patents filed by tenant incubatee.
- Number of NBEs, which are having sales revenue.
- The number of NBEs graduating or exiting from incubator within a given time frame, the number of NBEs which are merged or overtaken by other firm usually a corporate major and the number of firms which have been closed while being a tenant. Table 4 lists business incubation performance metrics (BIPMs) identified by various studies in this regard over time.

Table 4 Business Incubation Performance Metrics (BIPMs)

Article	BIPMs identified
Hackett and Dilts (2004b), Hamdani (2006), Udell (1990)	• Job growth by incubatee
Hackett and Dilts (2004b), Hamdani (2006)	• Financial performance of incubatee
Hackett and Dilts (2004b)	• Developmental advances of Incubatee at incubator exit time
Allen and Weinberg (1988), (Hamdani 2006)	• Ratio of 'number of firms exiting the incubator' to the 'number of firms discontinuing operations while still a tenant' • The number of firms graduating, merging, overtaken and closed while being a tenant
Hamdani (2006), Udell (1990)	• Number of tenant companies performing R&D and new patents filed • Incubatees having sales revenue growth
Hamdani (2006)	• The capital structure of tenant companies

References

Aerts, K., Matthyssens, P., & Vandenbempt, K. (2007). Critical role and screening practices of European business incubators. *Technovation, 27*(5), 254–267. https://doi.org/10.1016/j.technovation.2006.12.002.

Allen, D. N., & Weinberg, M, (1988). State investment in business incubators. *Public Administration Quarterly, 12*(2), 196–215.

Bergek, A., & Norrman, C. (2008). Incubator best practice: A framework. *Technovation, 28*(1–2), 20–28. https://doi.org/10.1016/j.technovation.2007.07.008.

Bøllingtoft, A., & Ulhøi, J. P. (2005). The networked business incubator—Leveraging entrepreneurial agency? *Journal of Business Venturing, 20*(2), 265–290. https://doi.org/10.1016/j.jbusvent.2003.12.005.

Bruneel, J., Ratinho, T., Clarysse, B., & Groen, A. (2012). The evolution of Business incubators: Comparing demand and supply of business incubation services across different incubator generations. *Technovation, 32*(2), 110–121. https://doi.org/10.1016/j.technovation.2011.11.003.

Campbell, C., Kendrick, R. C., & Samuelson, D. S. (1985). Stalking the latent entrepreneur: Business incubators and economic development. *Economic Development Review*, 43–48.

Carland, J. W., Hoy, F., & Carland, J. A. C. (1988). "Who is an entrepreneur?" Is a question worth asking. *American Journal of Small Business*.

Chan, K. F., & Lau, T. (2005). Assessing technology incubator programs in the science park: The good, the bad and the ugly. *Technovation, 25*(10), 1215–1228. https://doi.org/10.1016/j.technovation.2004.03.010.

Choi, Y. R., & Shepherd, D. A. (2004). Entrepreneurs' decisions to exploit opportunities. *Journal of Management, 30*(3), 377–395. https://doi.org/10.1016/j.jm.2003.04.002.

Crespi, G. (2012). Innovation and productivity: Evidence from six Latin American countries. *World Development, 40*(2), 273–290. https://doi.org/10.1016/j.worlddev.2011.07.010.

Díez-Vial, I., & Montoro-Sánchez, Á. (2016). How knowledge links with universities may foster innovation: The case of a science park. *Technovation, 50–51,* 41–52. https://doi.org/10.1016/j.technovation.2015.09.001.

European Commission Enterprise Directorate General. (2002). *Benchmarking of Business Incubators*. Brussels.

Gartner, W. B. (1989). "Who is an entrepreneur?" Is the wrong question. *Entrepreneurship: Theory & Practice*, *13*(4), 47–68. https://doi.org/10.1210/jc.2003-031037.

Grimaldi, R., & Grandi, A. (2005). Business incubators and new venture creation: An assessment of incubating models. *Technovation, 25*(2), 111–121. https://doi.org/10.1016/S0166-4972(03)00076-2.

Hackett, S. M., & Dilts, D. M. (2004a). Real options-driven theory of business incubation. *Journal of Technology Transfer, 29,* 41–54.

Hackett, S. M., & Dilts, D. M. (2004b). A systematic review of business incubation. *Journal of Technology Transfer*, *29*, 55–82. https://doi.org/10.1023/B:JOTT.0000011181.11952.0f.

Hackett, S. M., & Dilts, D. M. (2008). Inside the black box of business incubation: Study B—Scale assessment, model refinement, and incubation outcomes. *Journal of Technology Transfer, 33*(5), 439–471. https://doi.org/10.1007/s10961-007-9056-9.

Hamdani, D. (2006). Conceptualizing and measuring business incubation. *Cancer, 1*(S3), 1–25. https://doi.org/10.1002/1097-0142(19910201)67:3+%3c774::AID-CNCR2820671404%3e3.0.CO;2-I.

Hansen, M. T., Chesbrough, H. W., Nohria, N., & Sull, D. N. (2000). Networked incubators. Hothouses of the new economy. *Harvard Business Review*, *78*(5), 74–84, 1990

Katz, N., Lazer, D., Arrow, H., & Contractor, N. (2004). Network theory and small groups. *Small Group Research, 35*(3), 307–332. https://doi.org/10.1177/1046496404264941.

Kimatu, J. N. (2016). Evolution of strategic interactions from the triple to quad helix innovation models for sustainable development in the era of globalization. *Journal of Innovation and Entrepreneurship, 5*(1), 16. https://doi.org/10.1186/s13731-016-0044-x.

Lalkaka, R. (2001). "Best Practices" in business incubation: Lessons (yet to be) learned Rustam. *Human Development*, 1–35. https://doi.org/10.1111/j.1537-2995.2009.02262.x.

Lalkaka, R. (2002). Technology business incubators to help build an innovation-based economy. *Journal of Change Management, 3*(2), 167–176. https://doi.org/10.1080/714042533.

Lalkaka, R., & Abetti, P. A. (1999). Business incubation and enterprise support systems in restructuring countries. *Creativity & Innovation Management, 8*(3), 197. https://doi.org/10.1111/1467-8691.00137.

Löfsten, H., & Lindelöf, P. (2005). R&D networks and product innovation patterns—Academic and non-academic new technology-based firms on Science Parks. *Technovation, 25*(9), 1025–1037. https://doi.org/10.1016/j.technovation.2004.02.007.

Manimala, M. J., & Vijay, D. (2012). *Technology Business Incubators (TBIs): A perspective for the emerging economies*. Indian Institute of Management Bangalore.

McAdam, M., & McAdam, R. (2008). High tech start-ups in University Science Park incubators: The relationship between the start-up's lifecycle progression and use of the incubator's resources. *Technovation, 28*(5), 277–290. https://doi.org/10.1016/j.technovation.2007.07.012.

Merrifield, D. B. (1987). New business incubators. *Journal of Business Venturing*. https://doi.org/10.1016/0883-9026(87)90021-8.

Meyer, M. (2003). Academic entrepreneurs or entrepreneurial academics? Research-based ventures and public support mechanisms. *R&D Management, 33*(2), 107–115. https://doi.org/10.1111/1467-9310.00286.

Mian, S., Lamine, W., & Fayolle, A. (2016). Technology business incubation: An overview of the state of knowledge. *Technovation, 50–51,* 1–12. https://doi.org/10.1016/j.technovation.2016.02.005.

Montmartin, B., & Herrera, M. (2015). Internal and external effects of R&D subsidies and fiscal incentives: Empirical evidence using spatial dynamic panel models. *Research Policy, 44*(5), 1065–1079. https://doi.org/10.1016/j.respol.2014.11.013.

Moreira, A. C., Marta, F. S., & Carvalho, M. F. S. (2012). Incubation of new ideas: extending incubation models to less-favored regions. In T. Burger-Helmchen (Ed.), *Entrepreneurship, Creativity and Innovative Business Models* (pp. 41–58). https://doi.org/10.5772/36705.

Moreira, A. C., & Carvalho, M. F. S. (2015). Incubation of new ideas: Extending incubation models to less-favored regions (pp. 40–58). www.intechopen.com. https://doi.org/10.5772/36705.

Peters, L., Rice, M., & Sundararajan, M. (2004). The role of incubators in the entrepreneurial process. *The Journal of Technology Transfer, 29,* 83–91. https://doi.org/10.1023/B:JOTT.0000011182.82350.df.

Rice, M. P. (2002). Co-production of business assistance in business incubators: An exploratory study. *Journal of Business Venturing, 17*(2), 163–187. https://doi.org/10.1016/S0883-9026(00)00055-0.

Robinson, D. F. (2010). The co-evolution of business incubators and national incubator networks in emerging markets. *Journal of Technology Management and Innovation, 5*(3), 1–14. https://doi.org/10.4067/S0718-27242010000300001.

Rothaermel, F. T., & Thursby, M. (2005a). Incubator firm failure or graduation?: The role of university linkages. *Research Policy, 34*(7), 1076–1090. https://doi.org/10.1016/j.respol.2005.05.012.

Rothaermel, F. T., & Thursby, M. (2005b). University-incubator firm knowledge flows: Assessing their impact on incubator firm performance. *Research Policy, 34*(3), 305–320. https://doi.org/10.1016/j.respol.2004.11.006.

Rubin, T. H., Aas, T. H., & Stead, A. (2015). Knowledge flow in technological business incubators: Evidence from Australia and Israel. *Technovation, 41,* 11–24. https://doi.org/10.1016/j.technovation.2015.03.002.

Scillitoe, J. L., & Chakrabarti, A. K. (2010). The role of incubator interactions in assisting new ventures. *Technovation, 30*(3), 155–167. https://doi.org/10.1016/j.technovation.2009.12.002.

Smilor, R. (1987). Managing the incubator system: Critical success. *IEEE Transactions on Engineering Management*, *EM-34*(3), 146–155.

Smilor, R. W., Gibson, D. V., & Dietrich, G. B. (1990). University spin-out companies: Technology start-ups from UT-Austin. *Journal of Business Venturing, 5*(1), 63–76. https://doi.org/10.1016/0883-9026(90)90027-Q.

Soetanto, D., & Jack, S. (2016). The impact of university-based incubation support on the innovation strategy of academic spin-offs. *Technovation, 50–51,* 25–40. https://doi.org/10.1016/j.technovation.2015.11.001.

Theodorakopoulos, N., Kakabadse, N. K., & Mcgowan, C. (2014). What matters in business incubation? A literature review and a suggestion for situated theorising. *Journal of Small Business and Enterprise Development, 21*(4), 602–622. https://doi.org/10.1108/JSBED-09-2014-0152

Treibich, T., Konrad, K., & Truffer, B. (2013). A dynamic view on interactions between academic spin-offs and their parent organizations. *Technovation, 33*(12), 450–462. https://doi.org/10.1016/j.technovation.2013.06.012.

Udell G. G, (1990), Are business incubators really creating new jobs by creating new business and new products?. *Journal of Product Innovation Management, 7*(2), 108–122. https://doi.org/10.1016/0737-6782(90)90053-H

UNCTAD. (2011). A Framework for Science, Technology and Innovation Policy Reviews. United Nations publication. UNCTAD/DTL/STICT/2011/7.

Vij, S., & Jhanji, H. (2013). Business incubation: A review of research orientations, impacts and determinants of success. In *The Tenth Biennial Conference on Entrepreneurship, (EDI) Entrepreneurship Development Institute of India*, Ahmadabad, February 20–22, 2013 (pp. 1–13).

Von Zedtwitz, M. (2003). Classification and management of incubators: Aligning strategic objectives and competitive scope for new business facilitation. *International Journal of Entrepreneurship and Innovation Management, 3*(1–2), 176–196. https://doi.org/10.1504/IJEIM.2003.002227.

Von Zedtwitz, M., & Grimaldi, R. (2006). Are service profiles incubator-specific? Results from an empirical investigation in Italy. *Journal of Technology Transfer, 31,* 459–468.

Wiklund, J., & Shepherd, D. (2005). Entrepreneurial orientation and small business performance: A configurational approach. *Journal of Business Venturing, 20*(1), 71–91. https://doi.org/10.1016/j.jbusvent.2004.01.001.

Zhang, H., Wu, W., & Zhao, L. (2016). A study of knowledge supernetworks and network robustness in different business incubators. *Physica A: Statistical Mechanics and its Applications, 447,* 545–560. https://doi.org/10.1016/j.physa.2015.12.051.

The Neural Correlates of Decision-Making: Review and Research Agenda

Kirti Sharda

1 Introduction

Neuroeconomics combines classical economics and neuroscience to deepen our understanding about the role played by human brain in economic decision-making (Fehr et al. 2005; Goetz and James III 2008). Classical economics is the "science of choice, constrained by scarce resources and institutional structure" (Camerer 2013, p. 426) and neuroeconomics provides a mechanistic, mathematical and behavioural framework to understand these choices (Glimcher and Rustichini 2004).

The revealed preference approach (Sameulson 1938) which used observable choices to infer unobservable preferences (subject to certain axiomatic constraints) fundamentally changed the field of economics. Human behaviour was attributed to maximisation of some utility function which could be rigorously tested through empirical means. But the psychological and/or neurobiological accounts of preference were ignored for several decades.

Economic theorists were confounded by the emergence of numerous axioms aimed at explicating the same arrays of choices. A powerful way to adjudicate among theories was to insist that theorists must articulate specific predictions about underlying neural correlates (Glimcher and Fehr 2014). Herbert Simon's assertion that humans are boundedly rational (Simon 1955; 1997), Kahneman and Tversky's modelling of choices in real-life (1979), and Gigerenzer and Todd's (1999) findings on the trade-off between efficient choice, computational complexity and use of simple heuristics in decision-making highlighted that humans often failed to execute optimal courses of actions in their daily lives. In parallel, advances in neuroscience and cognitive psychology revealed the constraints of neural activity (in a way defining the boundaries of boundedly rational decision-making) and thus supported the development of neuroeconomics.

K. Sharda (✉)
Indian Institute of Management Ahmedabad, Ahmedabad, India
e-mail: kirti@iima.ac.in

P. Ray et al. (eds.), *Economics, Management and Sustainability*,
https://doi.org/10.1007/978-981-13-1894-8_13

The interest of economists in psychology and neurobiology dates back to late nineteenth century when Thorstein Veblen (1898) posited in his essay, *"Why is Economics Not an Evolutionary Science?"*, that economic behaviour can be understood by studying the underlying mechanisms which created those behaviours. There were also fantastic ideas, by earlier economists including Edgeworth, Fisher, and Ramsey (see Colander 2007), about developing "hedonimeters" to link biology and choices directly. However, we had to wait till late 1990s and early 2000s to see significant growth in the study of neuroeconomics.

Evolutionary theory posits that the singular goal of any behaviour of natural organisms is maximisation of inclusive fitness which maximises the survival of the organism's genetic code. Accordingly, the primary function of the nervous system, which mediates behaviour, is to produce motor responses, under conditions of uncertainty, that yield the highest possible inclusive fitness for an organism. Economic theory provides us with the optimal decision for a given context through a rational computational route. This optimum economic solution combined with the goal of nervous system to make decisions aligned with inclusive fitness could help us in understanding the spectrum of human behaviour (Glimcher 2003).

Neuroeconomics today informs research in a range of management disciplines including finance (Efremidze et al. 2017; Frydman et al. 2014), consumer behaviour (Berns and Moore 2012; Egidi et al. 2008), organisational behaviour (Beugré 2009; Lori 2017), strategy (Hodgkinson and Healey 2011), and others.

1.1 Anatomy and Essential Features of the Brain: Significance for Neuroeconomics

Detailed examination of the brain anatomy, from a neuroeconomic perspective, has centred around basal ganglia and cerebral cortex, which form part of the telencephalon or forebrain. While cerebral cortex is a much more recently evolved structure, the basal ganglia is more evolutionary ancient. Some of the sub-regions of the basal ganglia like the striatum which consists of the caudate and putamen,[1] the globus pallidus, substantia nigra pars reticulate[2] and the dopaminergic system[3] figure prominently in neuroeconomic studies. The amygdala[4] along with hypothalamus,[5] which are parts of the telencephalon, receive inputs from many sensory systems and in turn

[1]The ventral striatum plays an important role in encoding of option values during choice tasks.

[2]The core circuit of basal ganglia receives information from frontal cortex (through caudate and putamen), and post-processing, transfers it back to frontal cortex (through globus pallidus and substantia nigra pars reticulate).

[3]There is substantive evidence that the dopamine-releasing neurons in the dopaminergic system encode a reward prediction error signal which aids reinforcement learning.

[4]The amygdala has been implicated in a host of studies related to psychological states.

[5]The hippocampus figures repeatedly in neuroscientific studies of learning and memory.

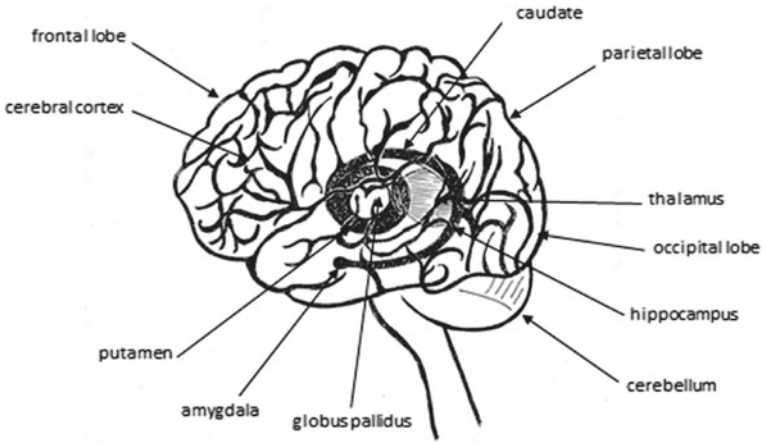


Fig. 1 Anatomy of the brain

intimate and regulate activity in several frontal cortical areas (Glimcher 2014a). A general anatomy of the brain, with relevant regions, is presented in Fig. 1.

Certain features of the brain stand out from a neuroeconomic perspective, namely modularity of the brain, neuronal stochasticity, synaptic plasticity and asymmetric information flow.

The behavioural and cognitive abilities of the brain are deemed to be the outcome of a complex multi-tiered processing system involving independent sub-processes which interconnect with other systems through well-defined inputs and outputs. Fodor (1983)[6] highlighted the presence of such independent modules, which appear to handle a certain class of information, perform specific computations and pass on their outputs to other modules for further processing (Glimcher 2014a).

The strong limits imposed on the firing rates of cortical neurons define the upper bounds on the amount of information that a neuron can transmit. The stochastic nature of neuronal firing connects the analysis of neuronal activity to random utility or discrete choice approach. Synaptic plasticity is the ability of synapses to strengthen or weaken over time. The biochemical mechanism by which information is stored in the nervous system over periods of days or longer is a process of synaptic modification (Glimcher 2014a).

Selective or asymmetric flow of information can lead to time lags in various parts of the brain receiving the same information. Evolutionary processes prefer conservation of energy and this imposes certain limitations on the brain which is the most energy-intensive organ in the human body. These limitations present themselves in the form of limited neural connectivity constraining free flow of information in the brain (Alonso et al. 2014).

[6]Fodor insisted that the modularity of brain should be considered only in the context of psychological studies. Nevertheless, his propositions gained significant attention in the neurobiological community.

1.2 Tools and Methods in Neuroeconomics

Techniques used by neuroeconomists to observe brain activity can be broadly classified into two categories: measurement techniques, which measure changes in brain function, while a subject engages in an experimental task, and manipulation techniques, which examine how disruptions in neural processing of specific regions change choices and related behaviour. Measurement techniques are "correlational", while manipulation techniques are deemed as "causal" in their approach (Ruff and Huettel 2014).

Neuroscientists typically select a research method based on requisite temporal resolution, spatial resolution and invasiveness (Ruff and Huettel 2014). The methods employed to identify the neural foundations of economic behaviours involve a wide range of techniques, such as neurophysiological measures, neuroimaging studies, neuromodulation (including neurostimulation), brain lesion analysis. The commonly used tools are fMRI, fNIRS, EEG, MEG, PET, TMS, tDCS, single-unit recording, pharmacological interventions, and brain lesion studies (See Crockett and Fehr 2014; Efremidze et al. 2017; Houser and McCabe 2014; Kable 2011; Lin et al. 2010; Ruff and Huettel 2014; Vercoe and Zak 2010; Volk and Köhler 2012; Zhao and Siau 2016). A brief summary of the primary tools and methods used in neuroeconomics has been presented in Exhibit (Table 1).

Other methods such as eye tracking (fixations and saccades), scanpaths, pupil size, blink rate, skin conductance rate, pulse rate and blood pressure have also been used to detect biological and emotional responses to stimuli in decision-making contexts (Zhao and Siau 2016).

Besides these, in non-human primates and rats, invasive methods such as microstimulation and optogenetics are used to detect and manipulate electrical activity in neurons with the help of inserted microelectrodes, intracranial light application or genetic engineering (Ruff and Huettel 2014).

Since different methods have unique strengths and weaknesses, current research, in the quest to understand the neurological underpinnings of human behaviour, makes an attempt to combine different techniques, either sequentially or in parallel, in order to achieve complementary support for its findings (Zhao and Siau 2016).

In this chapter, the primary research themes in neuroeconomic decision-making have been identified and organised around role of reinforcement learning systems in valuation and choice, value-based decision-making, decision-making under conditions of risk and ambiguity and intertemporal discounting. In addition to this, social preferences and context-dependencies in decision-making have also been discussed. Finally, fallacies in interpretations, methodological issues in neuroeconomics research, current concerns and future directions have been summarised.

Table 1 Tools and methods in neuroeconomics

Method	Objective	Type	Intervention	Description	Strengths	Limitations
Functional magnetic resonance imaging (fMRI)	Measurement	Non-invasive	Neurophysiological	Identification of neural activity by measuring changes in blood oxygenation levels (with the help of BOLD (blood-oxygenation-level-dependent) contrast signal	Non-invasive, convenient, superior spatial resolution leading to effective and accurate imaging in entire brain, including deeper regions. Moderate temporal resolution Easy access to equipment, trained personnel and experts	Expensive, potential safety risks, sensitive to noise emanating from equipment and small physiological variations
Functional near-infrared spectroscopy (fNIRS)	Measurement	Non-invasive	Neurophysiological	Similar to fMRI; Measurement of changes in cerebral blood oxygenation levels with the help of near-infrared light (from light sources placed on the scalp)	Non-invasive, small, portable, less obtrusive, less expensive, high temporal resolution	Poor spatial resolution, cannot measure activation throughout brain (limited to cortical activity less than 4 cm deep)
Electroencephalography (EEG)	Measurement	Non-invasive	Neurophysiological	Detection of electrical current (especially along the scalp) in order to measure voltage fluctuations resulting from ionic current within neurons	Relatively inexpensive, high temporal resolution	Imprecise spatial resolution
Magnetoencephalography (MEG)	Measurement	Non-invasive	Neurophysiological	Detection of magnetic fields produced by brain electrical currents with the help of magnetometers	Recording of data from entire brain including deeper regions. High spatial and temporal location	Expensive and not easily accessible. High maintenance and personnel costs Sensitive only to neurons oriented parallel to skull (not radially oriented)
Positron emission tomography (PET)	Measurement	Invasive	Metabolic neuroimaging	Measurement of blood flow in brain (as a proxy for neural activity) with the help of injected radioactive isotope (attached to a relevant neurotransmitter), and associated neurochemical changes	Provides information about different aspects of neural metabolism or neurotransmission. Can cover entire brain with moderate spatial resolution	Invasive, expensive, can be administered only in restricted doses and in tightly monitored samples. Limited temporal resolution
Transcranial magnetic stimulation (TMS)	Manipulation	Non-invasive	Neurostimulation	Stimulation of neurons through electromagnetic induction to generate electrical current and trigger action potentials	Non-invasive, high spatial resolution, high temporal resolution	Cannot access deep brain areas Sensitive to noise produced by procedure

(continued)

Table 1 (continued)

Method	Objective	Type	Intervention	Description	Strengths	Limitations
Transcranial direct current stimulation (tDCS)	Manipulation	Non-invasive	Neurostimulation	Use of low, constant, direct current to stimulate neurons in order to change spontaneous neuronal firing	Low noise emanating from procedure and experimental conditions. Relatively inexpensive. Can be administered to large sample sizes	Lower spatial resolution, low temporal precision
Pharmacological Interventions	Manipulation	Invasive	Neuromodulation	Manipulation of neuromodulator systems (neurotransmitters such as serotonin, dopamine, norepinephrine, and hormones such oxytocin and testosterone) by increasing or decreasing precursors, stimulating or blocking neuromodulator receptors with pharmacological agents, changing dosage of re-uptake inhibitors, oral or intravenous administration of neuro modulators, and neurotoxic lesions	Uniform administration to all participants. Provides evidence for role of neuromodulators in intertemporal choices, risk-based decision-making, reinforcement learning, social preferences, prosocial behaviours	Invasive, potential safety concerns
Brain lesion studies	Measurement[a]	Non-invasive	Neurophysiological	Systematic examination of behavioural disruptions that occur due to brain damage in patients. Also include single association studies (comparison between lesioned-patients and non-lesioned participants) and double association studies (comparison between patients grouped according to lesion sites)	Double dissociation studies provide strong empirical support for neural correlates of cognitions and behaviour	Difficult to locate comparable samples. Brain damage can have diffused and complex effects which are difficult to attribute to any one region or function
Single- recording	Measurement	Invasive	Neurophysiological	Measurement of changes in neuronal firing rate through insertion of microelectrode in brain	Reveals diversity of brain processes within a region which could be invisible to other methods	Invasive. Largely used in non-human subjects, with some exceptions. Laborious and expensive. Limited inferences

[a]Brain lesion studies are primarily measurement-oriented in humans. In animal subjects, they are also used as a manipulative technique

2 Role of Reinforcement Learning Systems in Valuation and Choice: Evidence from Neuroeconomics

Natural organisms are constantly challenged to optimise their behaviour in various environments. With the help of the computational frameworks provided by reinforcement learning (RL) systems, they rely on prior experiences to make predictions about the current context and choose appropriate behaviours which can lead to rewards or to avoidance of punishments. Since the computational algorithms of RL appear to have distinct neural correlates (such as the phasic activity of dopamine neurons) (Dayan and Niv 2008), neuroeconomists believe that an understanding of neural systems can contribute to the comprehension of choice behaviour. The neutrally distinct learning systems are detailed below.

2.1 Reinforcement Learning Systems

2.1.1 Pavlovian Learning

Pavlov (1927) discovered that when a stimulus is repeatedly paired with a reward or punishment, the stimulus by itself can trigger behaviour associated with the reward or create aversion for the punishment, in subsequent trials. As an example, if the reward is a food item, then the stimulus can elicit salivation, or in the case of pain as punishment the stimulus will elicit a withdrawal reflex. Pavlovian conditioning makes the organism learn to make predictions about the likely occurrence of significant events on the basis of preceding stimuli. Brain imaging studies have shown a strong association of several brain structures, particularly the amygdala, the ventral striatum and the orbitofrontal cortex, in this learning process (Gottfried et al. 2003; O'Doherty et al. 2002).

2.1.2 Instrumental Conditioning

Instrumental Conditioning associates actions with outcomes. These associations are shaped by reinforcement rules, and strengthen or weaken depending on the desirability of the outcome. The purpose is to make action choices such that they optimise goal achievement (Dayan and Niv 2008). Learning in instrumental systems is correlated with activity in the ventral striatum (Daw and O'Doherty 2014).

Habitual Learning Habitual Learning is a special case of instrumental learning where the learner repeats actions which have previously resulted in "satisfaction" without resorting to discrete evaluations with every new instance of the action. This implicit neural autopilot is activated for repeated low-value decisions but is incapable of analysing new contexts or re-evaluating previously experienced ones. Habitual

actions are also indifferent to short-term value changes and hence display short run choice elasticities of close to zero (Camerer 2013).

It has been demonstrated that increasing activity in right posterolateral striatum over the course of training relates to the emergence of habitual control (Tricomi et al. 2009). Diffusion tensor imaging has also shown significant correlation of the strength of connectivity between right posterolateral striatum and premotor cortex and the preference of habitual responding over goal-directed behaviour (de Wit et al. 2012).

The above three systems can be categorised as model-free RL where experiences are used to learn about values directly in order to arrive at optimal solutions without resorting to estimations or reference to a world model.

2.1.3 Goal-Directed or Model-Directed Learning

Goal-directed learning is classified as a model-based learning system, wherein experiences are used to construct a model which is a best approximation of the transitions and outcomes in the environment. Unlike model-free systems which require direct learning, goal values can be inferred from indirect methods like deliberation and communication. This method appears to be unique to humans (Daw and O'Doherty 2014). In a model-directed system, goal values associated with different choices are integrated with the help of abstract information, to arrive at an inclusive choice. The ventromedial prefrontal cortex (vmPFC), which has been of keen interest in the study of neuroeconomics due to numerous reports of correlation with expected value/utility, is associated with goal-directed learning (Balleine and O'Doherty 2010). Studies have also observed that expected value correlates in vmPFC comply more with model-based (versus model-free) values (Daw et al. 2011). Further, with the help of diffusion tenor imaging, it has been shown that goal-directed choice behaviour is correlated with the strength of interconnections between ventromedial prefrontal cortex and dorsomedial striatum (de Wit et al. 2012).

The above findings associate specific and segregated neural systems with each type of learning. However, these systems also interact among themselves, in a facilitatory or adversarial manner, to mediate the control of actions. For instance, the amygdala and ventral striatum have been implicated in human studies in Pavlovian-instrumental transfer (PIT); an example of a facilitatory interaction, where a Pavlovian association between a cue and reward can trigger an instrumental association between the reward and action (Prevost et al. 2012; Talmi et al. 2008). The ventral striatum has also been highlighted in scenarios where instrumental choice has been weakened by aversive Pavlovian cues (Chib et al. 2012). Further, lesion studies indicate that activity in the basolateral amygdala and ventral striatum is correlated with conditioned reinforcement effect (Cador et al. 1989). Similarly, neuropsychological models of addiction and compulsive behaviours have revealed a relationship between habitual and goal-directed systems, with over-active habitual systems overwhelming goal-directed choices and behaviour (Everitt and Robbins 2005).

2.2 Dopamine Reward Prediction Error

Prediction error, which denotes the difference between the predicted and the actual reward, is a key variable in learning systems. A large enough quantum of prediction error triggers more learning to update predictions. The more recent rewards get weighed more heavily compared to prior rewards, and the latter's weight declines exponentially with lag (Daw and Tobler 2014). According to Camerer (2013, p. 429), such reward values closely resemble an "economic construct of a stable utility for a choice".

- Studies have posited that dopamine plays a role in the *reward prediction error* of a particular event by signalling changes in the anticipated value of rewards. Dopamine neurons regulate their firing rates in sync with the reward prediction error with a positive surprise leading to a higher firing rate and vice versa, while rewards received on predicted lines do not lead to any change in firing rates (Bayer and Glimcher 2005; Nakahara et al. 2004). This implies that the dopamine neurons encode the reward prediction error. Studies have identified neurons which code the values of specific actions in dopaminergic target areas, that keep track of which actions have just been produced, and even pass on expectations about what rewards can be anticipated in the immediate future (Lau and Glimcher 2008). Further experiments have revealed that interventions which stimulate the dopaminergic system can influence the rate of learning (Pessiglione et al. 2006). These studies suggest that the dopaminergic system is used to imprint into the brain the subjective values of goods and actions which are learned from experience.

2.3 Future Directions

Learning systems play an important role in valuation and choice. However, how and why a particular system dominates behaviour at a certain point is still largely unanswered. Understanding is also limited about how neural regions interconnect with each other to compete or collaborate. More clarity is needed on the relationship between model-based and model-free reinforcement learning systems. While there has been some hypothesis to explain the competition between goal-directed and habitual behaviours, these have yet to be mapped to neural mechanisms. Similar associations need to be elucidated for Pavlovian vs. Instrumental competition (Daw and O'Doherty 2014).

Studies have shown when reward systems change, organisms do not "unlearn" previously learned predictive associations; they rather acquire a new association paradigm which might suppress the earlier relationship. There is lack of clarity on how this process fits in with the RL framework.

Current research in neuroeconomics has considered simple classical and instrumental conditioning cases. However, there are questions around how the dopaminergic system operates in more complicated scenarios like multiple actions contributing

to an outcome/outcomes, or in hierarchical tasks, which remain unanswered and provide interesting avenues for future research.

3 Value-Based Decision-Making

A biological framework, based on how millions of neurons interact in different brain systems, to understand choice behaviour is a major quest in neuroeconomic research. Value-based decision-making refers to a process in which choices are guided by subjective valuation of available options (Camerer 2013; Platt and Glimcher 1999; Rangel and Hare 2010). Two separate academic streams have emerged in decision-making research. Sensory neuroscientists (see Newsome, Britten, and Movshon 1989) study perceptual decision-making through the lens of sensation, while motor neuroscientists (see Glimcher 2003) approach the problem from the perspective of subjective valuation, that is idiosyncratic preferences or internal representations that guide choice. In this section, we follow the subjective valuation path and explore the basic principles and neural correlates of value-based decision-making. We also consider associated streams of research here such as cost-based decision-making and debates around utility measurement.

3.1 Valuation and Choice

Sequentially arranged neural mechanisms of valuation and choice combine together to create the value-based decision-making framework. The valuation mechanism learns, stores and retrieves the value of actions in consideration, and the choice mechanism using the output of the valuation circuit generates an actual choice from all available options (Glimcher 2014b). Valuation mechanism incorporates idiosyncratic preferences like those over risk and time through a variety of mechanisms including reinforcement learning processes. Choice mechanisms are focussed on choosing from the available choice set that element which has the highest value to us.

The Choice Circuit and its Neural Correlates Understanding the biological frameworks through which the neoclassical choice behaviour theory operates its argmax operation, to identify the highest valued option in the current choice set, has been a principal goal of neuroeconomic research over the last two decades. The current consensus says that subjective values are encoded through the topographical features of the brain and scalar qualities are encoded in neural firing rates. Stochasticity of neural firing rates could be responsible for the kind of variability in behaviour as described by the random utility theory models (see McFadden 2005) of economics.

Research has shown that signals closely related to subjective values have been detected from neurons in three interconnected topographic maps, namely the superior colliculus, frontal eye fields and the lateral intraparietal area (Ding and Hikosaka

2006; Platt and Glimcher 1999). Further research has also discovered that under simulated conditions neurons encode expected utility (reward magnitude/probability), not choice probability and further, in a stochastic manner which could account for unpredictability in behaviour. This lends support to both neuroscientific and economic explanations of decision variables (Dorris and Glimcher 2004). Also firing rates, which encode subjective value while organisms are considering their choice set, reconfigure to encode the choice once a decision has been made (Louie and Glimcher 2010).

The Valuation Circuit and its Neural Correlates Study of dopamine neurons and fMRI signals have discovered that activity in the striatum and the ventromedial prefrontal cortex consistently predicts people's preferences (Chib et al. 2009; Kable and Glimcher 2007; Levy and Glimcher 2011; Sanfey et al. 2003). Human preferences for various types of rewards are encoded in a single common neural currency. The subjective value signals generated by these areas always correlate with and predict choice behaviour. A complex network of brain areas contributes to the subjective value signals generated in the medial prefrontal cortex and the striatum. For instance, dorsolateral prefrontal cortex provides critical inputs for valuing social cooperation and goods that require or invoke self-control processes, while orbitofrontal cortex contributes to valuation of many consumable rewards. Neurons in amygdala play a critical role in the emotional regulation of value by generating contextual effects of fear and stress.

Many researchers support a multi-stage model which describes how decision-making is determined by subjective valuation signals (Levy and Glimcher 2012; Platt and Plassmann 2014). Accordingly, first varied features of choices are combined to create subjective valuation signals. The latter are transformed into action valuation signals. Finally, a comparison is made between the two valuation signals, and stored assessments are updated to improve future choices. These stages do not necessarily operate in sequence nor are fully separable.

3.2 Cost-Based Decision-Making

Neural mechanisms have been shown to play an active part in discriminating between costs associated with choice. These costs can be broadly divided into energy costs (that is, effort expended in pursuing a choice) and costs related to delay in rewards. The anterior cingulate cortex, the orbitofrontal cortex, and the striatum are the key brain areas associated with cost-based decision-making. There is evidence to show that effort and delay costs are handled by different neural mechanisms (Denk et al. 2005; Prevost et al. 2010; Wallis and Rushworth 2014). Model-based fMRI and computational modelling studies have also demonstrated that individuals are differentially sensitive to cognitive and physical effort with amygdala playing an important role in valuing rewards associated with cognitive effort.

3.3 Debates Around Utility Measurement

Measurement and comparison of reward utility for economic decision-making is complicated as the kind of utility can be as varied as experienced utility, decision utility, anticipated or predicted utility, and remembered utility (see Berridge and Aldridge 2008, for a taxonomy of reward utilities). Some clarity has emerged regarding the neural correlates of few types of utility, while confusion remains around others. For instance, research shows that the orbitofrontal cortex is robustly associated with experienced utility. But debates still continue about whether dopamine mediates a pure form of decision utility or that of remembered utility as a prediction error mechanism of reward learning (Bayer and Glimcher 2005; Berridge 2012; Glimcher 2011; Niv et al. 2012). The entanglement and conflation of distinct utility measures also leads to misrepresentations. For instance, decision utility which alludes to "wanting", and experienced utility which is synonymous with "liking", are frequently regarded in the same vein, though they are associated with unique neural systems and occur at distinctive points in time. Integration of diverse forms of utility remains a challenge given the multidimensionality of the problem (Witt and Binder 2013).

3.4 Future Directions

Theoretical models and empirical studies have begun to link traditionally divergent models of perceptual decision-making and valuation-based decision-making. While results of valuation circuits have been linked and generalised to discrete and simple choices, we are not yet sure whether these circuits can explain more abstract and complex choices. Further, even though these ubiquitous circuits have been generalised to the study of decision-making, we still do not know their general boundary conditions. Glimcher (2014b) recommends more investigations into models of valuation circuit and its interactions with choice circuit. Valuation circuits are riddled with redundancies, and it would be interesting to investigate whether these contribute to different stages in the decision-making process and thus represent different mental states (Platt and Plassmann 2014). Another interesting line of research will be to integrate the pharmacological results with those from neurophysiology to determine how neuromodulators such as dopamine affect neuronal encoding (Wallis and Rushworth 2014).

4 Intertemporal Choice

Intertemporal choices, choosing between outcomes occurring at different points in time, is associated with complex neural mechanisms, which encode proximate rewards and predict and evaluate distal rewards (Camerer 2013).

4.1 Intertemporal Discounting

Across species which have been tested, it has been noted that there is a consistent discounting of future rewards when compared with immediate rewards. Vast majority of studies have found that discounting of future outcomes tends to follow hyperbolic and quasi-hyperbolic functions rather than exponential functions (Ainslie and Haendel 1983; Soman et al. 2005). Similar results have been obtained while researching impulsive behaviour among addicts (Ainslie 1975).

fMRI experiments based on dual-systems hypothesis have observed that different sets of brain areas are activated while considering intertemporal choices. It has been observed that ventral striatum (VS), posterior cingulate cortex (PCC), and ventromedial prefrontal cortex (vmPFC) demonstrate higher activation while considering immediate reward choices when compared to choices involving only delayed rewards. Simultaneously, posterior parietal cortex (PPC) and lateral prefrontal cortex (LPFC) are activated to greater extent than other regions for a spectrum of choices across the board. Kable and Glimcher (2010) disputed the dual-system interpretation and contended that active brain regions might be reacting to larger subjective value of immediate rewards and not to the immediacy. Blood-oxygen-level-dependent (BOLD) activity in VS, PCC and vmPFC seems to support this contention as these regions show higher correlation with subjective values in comparison with objective reward dimensions (Kable and Glimcher 2007). While the role of LPFC and PPC is still disputed, multiple studies have affirmed that in intertemporal choice, VS, PCC, and vmPFC play a significant role (Ballard and Knutson 2009; Kable and Glimcher 2010; Pine et al. 2009). Peters and Büchel (2010) have suggested that a relationship exists between working memory, time preferences and patience, which adds salience to future rewards.

4.2 Self-control

Dual-system accounts of intertemporal choices are closely linked to notions of self-control. Thaler and Shefrin (1981) have postulated that a battle for internal control rages on between a myopic "doer" and a farsighted "planner". While the "doer" can be associated with the limbic system, the "planner" is identified with the prefrontal cortex. Hare et al. (2009) examined the neural circuitry associated with the planner-doer and self-control dimensions and found a strong correlation between dorsolateral prefrontal cortex (DLPFC) and self-control choices. DLPFC's significance in determining self-control has been corroborated by experiments where temporary disruption of the region by transcranial magnetic stimulation (TMS) led subjects to become more impatient. Left DLPFC has also been associated with a range of task-related functions involving working memory and inhibitory control (Camerer 2013; Peters and Büchel 2010).

4.3 Future Directions

Questions remain about the neural correlates of intertemporal discounting. It is possible that brain regions which get activated while appraising immediate rewards are different from those engaged during discounting calculations (Harrison 2008). Considerable research efforts are being directed towards resolving the debate between dual-system and unified system of discounting. While small neural units could be indifferent to subjective values, it is possible that the sum of activity across these units could lead to prominent BOLD activity across a wider brain region. This hypothesis remains open for examination. Further, within the same brain region, different areas show different rates of discounting (Tanaka et al. 2004), which calls for future examination.

Harrison (2008) also emphasizes the need for more precise operationalisation of variables studied in intertemporal discounting, such as front end delay on earlier options, as related confounding leads to competing explanations for apparently huge discount rates.

Demographic and socio-economic factors also seem to play role in intertemporal choice with Westerners displaying larger time-discount rates than Easterners. Future studies can investigate the interactions between biological, cultural and socio-economic factors in intertemporal choice (see Read and Read 2004; Takahashi et al. 2010).

5 Risky Choice

Choice under risk and ambiguity constitutes a distinctive form of decision-making as choice options yield multiple outcomes with varying probabilities (Camerer 2013). In this section, the neural processes and mechanisms involved in choice under risk and ambiguity are examined.

Regardless of the uncertainty, choice options have to be assigned values for comparison and selection. Value assignments can be either outcomes and probabilities or standard statistical measures of probability distribution like mean, variance and skewness. Brain areas engaged in value encoding like dopamine neurons, striatum, orbitofrontal cortex and medial prefrontal cortex have been associated with processing of probability and outcome information (Tobler and Weber 2014).

5.1 Choice Under Risk, Ambiguity, Gains and Losses

Ambiguity is characterised by lack of knowledge of outcome distribution, so a gamble with unknown pay-offs is ambiguous, while a gamble with well-defined pay-offs and probability distributions is termed risky (Smith et al. 2002). Numerous studies and

experiments have shown that humans avoid ambiguity in both gains and losses. We prefer the riskier gambles only in loss situations while avoiding them in gain situation (Tversky and Kahneman 1992; Smith et al. 2002).

Pay-off structure (gain/loss) is held separable, if not independent, from belief structure (ambiguity/risk), in economics and decision theory. Research by Smith et al. (2002) has shown brain activity in neocortical dorsomedial system while processing loss information in risk-based gambles, and in ventromedial region under other conditions of risk-gain and ambiguous-gain or ambiguous-loss. The researchers have postulated from this findings that brain areas underlying the belief and pay-off structures, while being dissociable, are functionally integrated and interact with each other.

Breiter et al. (2001) suggested brain structures, distinct from Smith et al. (2002), associated with decision-making and expectancy; these included amygdala, hippocampus, nucleus accumbens, orbitofrontal cortex, ventral tegmentum and sublenticular extended amygdala. Ambiguity, as compared to risk, displays increased BOLD activity in orbitofrontal cortex (Hsu et al. 2005; Levy et al. 2010), amygdala (Hsu et al. 2005) and in some studies also in parietal cortex (Bach et al. 2011), and this higher level of activity could be signalling that information is missing (Tobler and Weber 2014).

Experiments have shown that subjective value of risky and ambiguous choice options is commonly coded in the medial prefrontal cortex, posterior cingulate cortex and the striatum (Levy et al. 2010). It is also possible that individual regions, while being engaged in representing decomposed components of risky choice models, interact with common regions for valuation of choice options (Tobler and Weber 2014). Research also shows that individual differences in choice behaviour can be predicted based on relative activation of the brain areas identified above (Huettel et al. 2006).

5.1.1 Loss Aversion

Tom, Fox, Trepel, and Poldrack's (2007) study showed strong association between differences in value-related neural activity and degree of loss aversion inferred behaviourally. Studies by Yacubian et al. (2006) highlighted neural activity related to gains in ventral striatum, while temporal lobe regions (lateral to striatum) and amygdala showed activity related to losses. Camerer (2013) and De Martino et al. (2010) demonstrated that patients with inhibited activity in bilateral amygdala due to lesions did not exhibit any loss aversion. Some researchers used diverse physiological methods to examine these associations. For instance, Hochman and Yechiam (2011) demonstrated that the increase in heart rate and pupil dilation correlated significantly with losses relative to comparable gains, while Sokol-Hessner et al. (2009) showed (with the help of skin conduction response) that "perspective-taking" helps in diminishing loss-aversive behaviour.

5.1.2 Endowment Effect

Neuroeconomic research of endowment effect[7] is of recent vintage and is still primarily focussed on identifying basic neural correlates. Weber et al.'s (2007) subjects demonstrated higher activation in amygdala during the selling trials of digital copies of songs, relative to buying trials. Kahneman et al. (1990), in order to adjust for the relative wealth positions of the seller and buyer in an endowment effect scenario, offered a "choice price" option.[8] fMRI studies conducted by Knutson et al. (2008), using high-value consumer goods, have found neural activity to indicate that selling prices were greater than choice prices, which in turn were greater than buying prices. The medial prefrontal cortex activation, which is known to have a negative correlation with losses and a positive correlation with gains (see Knutson et al. 2003), in this experiment showed consistent negative correlation with buying and choice prices and positive correlation with selling prices. Further, negative correlation with buying prices was much stronger than with choice prices.

Knutson et al. (2008) found that activation in right insula was positively and significantly correlated with endowment effect estimates. Since activation of insula is closely identified with distress (Masten et al. 2009; Sanfey et al. 2003), these results support the hypothesis that quantum of distress experienced by participants is positively correlated with endowment effect. However, subsequent studies by Tom et al. (2007) could not find any evidence of neural activity within insula or amygdala which are supposed to be related to negative emotions. Rather they found a direct correlation, positive for gain and negative for losses, in the ventromedial prefrontal cortex and dorsal and ventral striatum. Additionally, asymmetric neural activity for higher quantum of reduction for losses compared to quantum of increase in profits was observed as expected with behavioural loss aversion. Some of these results have been questioned given the limitations of fMRI (Knutson and Greer 2008). A strictly behavioural study comparing loss aversion behaviour of participants with inhibited amygdala, who could not process fear, and normal participants offered further contradictory evidence (De Martino et al. 2010). While non-lesioned subjects exhibited standard loss-aversive behaviour, subjects with lesions in amygdala did not exhibit any loss aversion. A supplemental study found that the differences in loss aversion were limited only to loss situations and were not exhibited in gain situations.

[7]Endowment effect refers to the difference between the minimum amount a person is prepared to receive in order to give away something she/he owns (selling price), and the maximum amount a person is prepared to give in order to acquire the same (buying price).

[8]Choice price refers to the indifference price point where subjects are neutral between getting goods and an equivalent amount of money.

5.2 Statistical Moments and Value Assignments

Studies have suggested that varied brain regions allow for coding of statistical measures like mean and variance of rewards (Platt and Huettel 2008). The statistical moments of reward distributions can be weighed and integrated to create choice values, and choice behaviour can be thereby examined (Camerer 2013; Tobler and Weber 2014). But objective valuations of multiple statistical measure decompositions are scarce due to limitations in experimental design. However, electrophysiological and neuroimaging of single cells using specialised designs have shown that mean-variance decompositions appear to be implemented in the orbitofrontal cortex while insula appears to decompose risk processing by separate representations of variance and skewness risk (Tobler and Weber 2014).

5.3 Future Directions

Understanding of the neural mechanisms underlying choice under risk and ambiguity is still at fairly rudimentary levels. It is still not clear whether single or multiple neural systems are engaged simultaneously or whether and how measures of probability and magnitude are decomposed at the neural level. Dissociations reported in the fMRI domain still await confirmatory evidence from other methods.

6 Context Effects

Standard deterministic models of normative choice, like expected utility theory, assume that choices are largely independent of context. However, numerous studies have shown that a number of contextual factors, like size of choice set, framing of decision problem, temporal history, play a critical role in the decision-making process (Louie and De Martino 2014). Decision-making is a dynamic process subject to the spatial and temporal context of the choice itself.

Contextual effects can occupy a wide spectrum including emotional states, motivational states, environmental situations and the like (Buehler et al. 2007; Witt and Binder 2013). Mental states include cognitive overload, attention levels, physiological states such as hunger, exhaustion, pain, discomfort and emotions like fear or anxiety. These states could act as constraints or sources of helpful information. For instance, fear could serve as a useful signal for immediate danger or noise. Fear and anxiety can also force a bias for quick resolution of uncertainty (Camerer 2013; Caplin and Leahy 2001). Contextual factors also include deprivation states such as sleep deprivation which leads to slow, noisy decisions (Camerer 2013; Menz et al. 2012). High states of deprivation can lead to “hot” visceral conditions where physiological factors can have an overriding influence on future estimation. In such con-

ditions, future visceral states are not properly accounted for leading to discrepancies between different kinds of utility (Witt and Binder 2013).

6.1 Neural Correlates of Context-Dependent Valuation

Neurobiological constraints on neural activity, which defines the minimum and maximum activity levels, could be the underlying source of contextual effects (Louie and De Martino 2014). Influence of contextual factors has been observed in ventromedial prefrontal cortex, a region that is also associated with computation of goal values (Plassmann et al. 2008) and neural coding of prediction error signal (Nieuwenhuis et al. 2005).

De Martino et al. (2006) postulated that framing effects in decision-making is moderated by the emotional system through an affect heuristic. This was based on their discovery of significant activation of bilateral amygdala when a typical choice is made. But it was not conclusive whether the amygdala activity is associated with a choice input, or moderation of emotional system during the choice process, or post-choice. They also discovered a strong correlation between of framing effects and ventromedial prefrontal cortex (vmPFC) activation and concluded that vmPFC plays a prominent role in evaluating and integrating emotional and cognitive information during choice behaviour.

Sokol-Hessner et al. (2013) experiment using the reverse contrast method yielded more clarity and showed heightened activity in the insula, anterior cingulate cortex and dorsolateral prefrontal cortex. These areas are also implicated in conflict resolution, emotion regulation and response inhibition. Insula is an active participant of encoding bodily sensations (especially discomfort) which is consistent with the feelings of discomfort experienced while making risky decisions.

Recent studies have suggested that framing evokes different kinds of affect and overriding of this affect leads to limiting the framing effect (Miu and Crişan 2011). Susceptibility of individuals to framing effect has shown significant variability. Evidence suggests that the medial orbitofrontal cortex (mOFC) plays a key role in controlling the framing effect, by modulating the amygdala approach avoidance signal, leading to more consistent and context independent decisions (De Martino et al. 2006). mOFC lesions in macaques result in strong framing effects in choice behaviour (Noonan et al. 2010). It has also been shown that amygdala's susceptibility to framing effect can be modulated by a specific polymorphism of the serotonin transporter gene (SERT). This has been linked to a reduced functional connectivity between mOFC and amygdala, essentially indicating that participants carrying the SS polymorphism carry a reduced ability to counteract the biasing influence (Hariri et al. 2002; Louie and De Martino 2014).

Stress (even artificially induced) can increase risk aversion (see Porcelli and Delgado 2009). Viewing of negative emotion-laden images prior to choice-making influences risk aversion (Kuhnen and Knutson 2011). Stimulation of dorsolateral pre-

frontal cortex leads to increased risk aversion (Fecteau et al. 2007), while disruption of the same leads to decreased risk aversion (Knoch et al. 2006a).

However, context-dependency also introduces a fundamental ambiguity that there is no correspondence between value quantities and firing rates as the given value could be influenced by a variety of contextual factors. In fact, it is still not clear whether and how contextual neural coding underlies context-dependency at the behavioural level (Louie and De Martino 2014).

6.2 Do Emotions Play a Role in Decision-Making? Neural Evidence

Emotions play a significant role in decision-making and are integral to success. The emotional content of heuristic decision-making increases with the complexity of decisions (Forgas 1995; Lo and Repin 2002). This has been neurologically substantiated with lesion studies on patients with impaired ventromedial cortex. The impairment "degrades the speed of deliberation and also degrades the adequacy of choice" (Bechara and Damasio 2005, p. 339). Emotions like anticipated disappointment and regret act as learning signals influencing future decisions (Coricelli et al. 2007; Steiner and Redish 2014). It has also been observed that dread of an anticipated negative experience is linked to neural activity related to physical pain (Dayan and Seymour 2009).

Even though influence of emotions in decision-making is undisputed, relatively few studies have measured or manipulated emotional variables during decision-making (Lempert and Phelps 2014). Some of the difficulties in studying emotions relate to the neurological anatomy. The ventromedial cortex can be divided into caudal/posterior and rostral/anterior areas. While the former exhibit a direct connection with regions moderating emotions, the latter are indirectly connected (Ongur and Price 2000). The caudal/posterior areas process events with high probability and the rostral/anterior areas process the less probable outcomes. Given the indirect connections, outcomes with lower probability would require to be more emotionally intense, in order to produce a significant emotional reaction (Bechara and Damasio 2005; Goetz and James III 2008). This creates confounding problems.

Emotion modulates decision-making through two paths: incidental affect or through inclusion into the value computation process. Research has identified distinct affective processes that incidentally influence decisions (Lempert and Phelps 2014). For instance, stress impacts the prefrontal cortex function whereby the decision process becomes more habitual and automatic. As mentioned earlier, in affect-as-information model (Schwarz and Clore 1983), the affect acts as additional information, even when it is irrelevant, which influences the judgement or decision. Consequently, even subliminal manipulations of emotions have an impact on valuation of the choice set (Lempert and Phelps 2014).

6.3 Future Directions

Reference points, used to evaluate gains and losses in a relative manner, are a construct which is widely used to explain contextual effects, but little is known about the actual neurobiology underlying the computation of reference points. Flexibility of emotions, adaptive in one circumstance and maladaptive in another, leading to flexibility in decision-making is also relatively unexplored in neuroeconomic research. There is still very limited information about explicit linking of the neural circuits and emotions with value representations (Lempert and Phelps 2014). Impact of other factors, like drugs and cognitive overloads, which modulate emotions and consequently moderate behavioural loss aversion is also underexplored (Paulus et al. 2005).

7 Examining Social Preferences: Game Theory, Empathy and Theory of Mind

Game theory examines how individual players make decisions in a multi-player environment where decisions of one player can impact the opportunities and pay-offs of other players. This is an interesting research area for neuroeconomics as game theory links individual decisions to group level outcomes through precise structures (Houser and McCabe 2014). Neuroeconomic studies can help answer some of the key problems in social decision-making such as examining the incentives for individual decisions, influence of emotions in decision-making or how behaviour is a function of the bounded rationality of participants (Fehr et al. 2005).

7.1 Social Preferences in Decision-Making: Neural Evidence

7.1.1 Competition and Cooperation

Studies have shown that humans prefer cooperation as opposed to defection in social dilemmas, notwithstanding equivalent monetary gains, thereby indicating a desire for benefits beyond monetary gains (Fehr et al. 2005; Fehr and Camerer 2007). Neuroimaging studies have revealed that dorsal striatum is activated in cooperative scenarios (Rilling et al. 2002), while ventromedial prefrontal cortex and ventral striatum are activated in competitive scenarios (Cikara et al. 2011; Dvash et al. 2010). Interestingly, the latter also figure dominantly in the "reward"-related neural networks.

Votinov et al. (2015) examined neural activity in strictly competitive games where one player's outcome is negatively correlated with the outcome of her/his opponent. They discovered two distinct neural activation areas associated with the two types of winning, one by direct increase in pay-offs (in ventromedial prefrontal cortex,

typically associated with rewards) and the other by loss avoidance (in precuneus and temporoparietal junction, typically associated with mentalising and empathy). Both types of winnings also showed activation in striatum.

Reputation plays an important role in repeated games with private information. Studies by Bhatt et al. (2012) found direct correlation of amygdala activity (associated with threat and risk perception) with doubts of credibility of the other player. Evidence was also found that beliefs were updated with data from recent behaviour in a continuous process. Separate but interconnected neural structures appear to mediate the uncertainty of the other players' behavioural reputation. Delgado et al. (2005) observed differential neural activation in caudate nucleus and several other areas in response to cooperative and competitive behaviour by opponents. Interestingly difference in activity was not observed where the other player had a good reputation to start with. This resonates with the view that subsequent "bad" behaviour is excused when the player has been "good" initially (Camerer and Hare 2014).

Lambert et al. (2017) investigated the role of neuromodulators (oxytocin) in improving the accuracy of social decision-making in complex situations. In an assurance game, simulating a win–win environment, oxytocin increased nucleus accumbens activity and seemed to facilitate cooperation. In a chicken game, simulating a win–lose environment where aggression while desirable could be fatal if the partner also responds with aggression, oxytocin down-regulated the amygdala and increased the valence of cues to decide on aggression or retreat moves.

7.1.2 Inequity Aversion and Reciprocity in Social Dilemmas

Sanfey et al.'s (2003) fMRI study demonstrated that insula activity is positively correlated with propensity to reject unfair offers in an Ultimatum Game. The activation of insula (and its correlation with negative affect) indicates that negative emotions drive rejection decisions in Ultimatum Game. Knoch et al. (2006b) found evidence for the crucial role of right DLPFC in mediating between self-interest and fairness impulses. In this study, the right DLPFC, implicated in overriding or weakening self-interest impulses, was inhibited with repetitive transcranial magnetic stimulation (rTMS) and subjects were found to be more willing to act selfishly. Moreover, where the opponent was a computer, which tends to inhibit the reciprocity motive, the right DLPFC disruption showed no behavioural effect.

DeQuervain et al. (2004) found evidence that people have a preference for punishing norm violations. Neural activity in a two-player sequential social dilemma game, where the players had the opportunity to punish the partner for abuse of trust, was observed using positron emission tomography (PET) imaging. It was noted that subjects experienced the punishing as satisfactory or rewarding and that the quantum of anticipated satisfaction and level of punishment was positively correlated (Fehr et al. 2005).

7.1.3 Trust

While oxytocin is associated with trust levels, it does not seem to change the evaluation of the other player's trustworthiness. It seems to operate more at the level of subject's preferences by making the subject more optimistic and inhibiting their exploitation aversion making them more receptive to the risk of being exploited (Bohnet and Zeckhauser 2004; Fehr et al. 2005; Kosfeld et al. 2005).

Krueger et al. (2007) indicated that two different trust mechanisms are in play in a repeated two-player trust game. The initial mechanism involving anterior paracingulate cortex makes way for spatial activation in the later stages. This suggests the operation of a conditional trust system in the later stages of the game, while avoiding trust in the early stages when temptations to defect are high.

7.1.4 Attitude Towards Strategic Uncertainty

The uncertainty of another person's actions in a cooperation task is referred to as strategic uncertainty (Ekins et al. 2013). Chark and Chew (2015) used a combined behavioural and neuroimaging approach to study response to strategic uncertainty as compared to non-strategic uncertainty. Their findings indicate that subjects showed aversion to strategic ambiguity in competitive environments and were ambiguity seeking in cooperative environments. They also exhibited source preference (self-regarding) rather than social preference (other-regarding) in their valuation of expected utility.

7.2 Association with Theory of Mind and Empathy

Decision-making in a game theory situation presupposes that players build a theory of mind or engage in mentalising about other players which helps to understand their motivations, preferences and thereby predict their actions (McCabe and Singer 2008; Singer and Fehr 2005). While theory of mind is a cognitive understanding of another person's mental state, empathy refers to the "ability to share the feelings and affective states of others" (Singer and Tusche 2014; p. 514). Comparative study of brain areas involved when humans play against each other, and when humans play with computers, has shown involvement of medial prefrontal lobe in mentalising. Empathy-related activation has been observed in anterior cingulate cortex and anterior insula. Empathy studies have shown that people tend to positively value pay-offs when the other player has played fairly and negatively value the other player's pay-off during unfair play. This indicates that people prefer to punish unfair opponents and choose to cooperate with fair ones (Fehr and Gächter 2000).

7.3 Pharmacological Manipulations of Social Preferences

Pharmacological manipulations have highlighted some causal effects. For instance, increased testosterone drives more fair offers (Eisenegger et al. 2010); increased serotonin and benzodiazepine lead to reduced rejection rates (Crockett and Fehr 2014; Gospic et al. 2011). Prosocial behaviour and accuracy of higher-order decision-making are enhanced by oxytocin administration (Kosfeld et al. 2005; Lambert et al. 2017).

7.4 Future Directions

Bhatt and Camerer's (2005) and Mohr et al. (2010) studies establish a link between differential insula activity and first- and second-order beliefs. This association can support future research on self-referential bias in second-order beliefs and in risk-related decision-making.

In-depth research is required to understand the neurological factors which modulate empathetic brain responses to better understand the conditions under which prosocial and anti-social behaviour is fostered (Singer and Tusche 2014). More detailed examination of the theory of mind and its neural correlates could provide linkages between strategic thinking and neural activity (Camerer and Hare 2014). Understanding the neural basis of exceptional skills in bargaining which contribute to game play success, can aid in building valuable management skills and competencies such as strategic thinking and negotiation. Explorations of linkages between different brain regions can bring more specificity and depth to our understanding of the relationship between relevant neural circuitry and social decision-making.

8 Discussion and Future Research Directions

Neuroeconomics started gaining traction with the rise of the "neuroessentialism" view (Racine et al. 2005) that the definitive way of explaining human psychological experience can be through understanding the brain and its activity. This has been fostered by rising capabilities and accessibility of neuroimaging techniques like fMRI. With well-developed experimental designs researchers can explore the neural substrates of behaviour and understand the mental processes which lead to certain behaviour. This helps in developing better theories and also in testing theories in multiple ways (Becker et al. 2011; Volk and Köhler 2012).

In this section, we discuss the fallacies researchers need to guard against, methodological issues in neuroeconomics research, concerns and future directions.

8.1 Frequent Fallacies

"Forward inference" involves manipulating a specific psychological function and identifying its localised effects in brain activity (Huettel and Payne 2009). However, some researchers engage in the "reverse inference" while interpreting experimental results. Reverse inference involves fallacious analysis as the concerned researcher selects activated brain regions and makes inferences about the mental processes associated with it. This results in low predictive power as specific brain regions can be related with multiple cognitive processes (Poldrack 2006). While reverse inference may be used to generate novel hypotheses about potential neural correlates (Poldrack 2011), researchers need to be circumspect in its use and design appropriate boundaries to enhance specificity of neural studies (Hutzler 2014; Volk and Köhler 2012). The other common fallacy to guard against is mereological fallacy which ascribes psychological processes like memory, perception, thinking, imagery, belief, consciousness to the brain, while they might be emerging from interaction of multiple parts of the body (Bennett and Hacker 2003; Powell 2011).

8.2 Methodological Issues

Researchers have highlighted some methodological concerns regarding design, execution and statistical analysis of neuroeconomics experiments. A brief summary of the same has been shared here.

To begin with, statistical analysis of neural data can often be challenging due to very small sample sizes (Fumagalli 2014; Ortmann 2008), resulting in "data sets in which a few brains contribute many observations at each point in time, and in a time-series" (Harrison 2008; p. 312). Neuroeconomic predictions are valid only over very limited time intervals (spanning few hundred milliseconds) making it irrelevant both for real-world use and for other economists (Fumagalli 2014).

Experimental results can be contaminated through experimenter-expectancy or demand characteristics effects as in neuroeconomics laboratories researchers, not blind to research questions and hypotheses, may provide instructions to subjects (Ortmann 2008). Other factors which reduce signal-to-noise ratio of the research findings include artificiality of laboratory settings (Ortmann 2008) and the use of deception in experiments (see Coricelli et al. 2005; Knoch et al. 2006a; Rilling et al. 2002; Sanfey et al. 2003). Issues relating to representativeness of stimuli which figured in behavioural economics apply to current neuroeconomics experiments as well (Ortmann 2008).

Resolution capabilities of the available imaging techniques are not able to capture the physiological heterogeneity of many neural structures and circuits, limiting our understanding of the complex functions of these structures (Fox and Poldrack 2009). Further, invasive data-collection techniques like repetitive transcranial mag-

netic stimulation can lead to unknown, irreversible damage in experimental subjects (Ortmann 2008).

Socio-demographic diversity is often lacking in sample sets which makes creating consensus on the applicability of results across contexts challenging (Harrison 2008). Further social, cognitive and affective capacities change over the course of life due to different developmental trajectories of the respective underlying neural structures. Initial studies have shown the potential benefits of investigating the developmental aspects of neuroeconomics (Steinbeis et al. 2012; Singer and Tusche 2014).

Concerns have been raised that researchers are driven more by the availability of new technology rather than by deeper understating of the underlying brain function or of what they are recording. As a result, when data are being collected or analysed, errors may go undetected, leading to inaccurate results. Research results may lead to overstated or implausible claims or may be reinterpreted to fit a previously held view (Cabeza and Nyberg 2000; Harrison 2008; Ortmann 2008; Ruff and Huettel 2014).

Triangulation through multiple research methods like pharmacological interventions, cross-cultural designs and genetic-imaging approaches could be helpful in arriving at a holistic comprehension of neural interconnections and can lead to a better understanding of the underlying affective, motivational and cognitive processes in decision-making (Singer and Tusche 2014; Volk and Köhler 2012).

8.3 Neuroeconomic Research: Concerns and Future Directions

Technology available currently imposes limitations on our understanding of the brain's neurobiological complexity. For example, we are not able to capture the structural features of many brain regions at a granular level nor are we able to grasp the extensive heterogeneity which resides at the cellular level (Fox and Poldrack 2009).

Neuroeconomics has identified the neurological underpinnings of many behaviours examined through prospect theory and provides strong evidence that "anomalies" are real. Prospect theory interprets risk attitudes in multiple ways like loss aversion, diminishing sensitivity to money and probability. During empirical neuroeconomic research, we run the risk of conflating these factors. We also need to establish whether an activity is causal and necessary for the presence of the phenomena and is not just an effect of these factors. This calls for a combined approach of identifying the neural correlates and establishing their necessity for the presence of a phenomenon (Fox and Poldrack 2009).

Some researchers have questioned the appropriateness and relevance of game theory contexts in understanding the reality of social preferences (see Binmore 2007; Levitt and List 2007). Use of subjective labels such as "trust" and "trustworthiness" to describe the behaviour of participants in laboratory experiments can lead to misperception around motives and can lead to situations where labelling is confused with

explanation. Precise operationalisation of variables is needed to control for possible confounds (Harrison 2008; Rubinstein 2006).

Context-dependencies are difficult to untangle given their multiplicity and simultaneity. Huettel and Payne (2009) urge researchers to be cautious in their interpretations, given the multidimensionality of context effects, and to be circumspect in providing overgeneralised explanations of results.

One key concern around neuroeconomic research is whether the validity of results obtained from simple games and choices can be extended to more complicated contexts (Ortmann 2008). While it is commonly believed that results obtained in a laboratory setting may not find resonance in field, Volk and Köhler (2012) highlight the availability of strong empirical evidence that majority of laboratory results demonstrate generalizability in "real world".

9 Conclusion

Neuroeconomics has managed to grow beyond the initial scepticism: whether it will add any value beyond correcting the "errors" believed to pervade economics and whether it will provide better answers to traditional questions. Real-world behaviour of economic agents in decision-making situations often tends to be suboptimal and deviates from rational optimal behaviour. Study of neural circuitry helps us comprehend the computations an agent makes, and how those computations are made. This knowledge might help us in understanding the deviations from optimal behaviour in a more fundamental way. Paradoxically, as Glimcher (2003) states, the value of neuroeconomic research might be enhanced by studying precisely those events when economic agents do not behave in the optimal manner, as predicted by theory.

References

Ainslie, G. (1975). Specious reward: A behavioral theory of impulsiveness and impulse control. *Psychological Bulletin, 82*(4), 463–496.

Ainslie, G., & Haendel, V. (1983). The motives of the will. In E. Gottheil, K. A. Druley, T. E. Skoloda, & H. M. Waxman (Eds.), *Etiologic aspects of alcohol and drug abuse* (pp. 119–140). Springfield, IL: Thomas.

Alonso, R., Brocas, I., & Carrillo, J. D. (2014). Resource allocation in the brain. *Review of Economic Studies, 81,* 501–534.

Bach, D. R., Hulme, O., Penny, W. D., & Dolan, R. J. (2011). The known unknowns: neural representation of second-order uncertainty, and ambiguity. *Journal of Neuroscience, 31,* 4811–4820.

Ballard, K., & Knutson, B. (2009). Dissociable neural representations of future reward magnitude and delay during temporal discounting. *Neuroimage, 45*(1), 143–150.

Balleine, B. W., & O'Doherty, J. P. (2010). Human and rodent homologies in action control: corticostriatal determinants of goal-directed and habitual action. *Neuropsychopharmacology, 35,* 48–69.

Bayer, H. M. & Glimcher, P. W. (2005). Midbrain dopamine neurons encode a quantitative reward prediction error signal. *Neuron, 47*(1), 129–141.

Becker, W. J., Cropanzano, R., & Sanfey, A. G. (2011). Organizational neuroscience: Taking organizational theory inside the neural black box. *Journal of Management, 37*, 933–961.

Bennett, M. R., & Hacker, P. M. S. (2003). *Philosophical foundations of neuroscience*. Malden, MA: Blackwell Publishing.

Berns, G. S., & Moore, S. E. (2012). A neural predictor of cultural popularity. *Journal of Consumer Psychology, 22,* 154–160.

Berridge, K. C. (2012). From prediction error to incentive salience: mesolimbic computation of reward motivation. *European Journal of Neuroscience, 35*(7), 1124–1143.

Berridge, K. C., & Aldridge, J. W. (2008). Decision utility, the brain and pursuit of hedonic goals. *Social Cognition, 26*(5), 621–646.

Bhatt, M., & Camerer, C. F. (2005). Self-referential thinking and equilibrium as states of mind in games: FMRI evidence. *Games and Economic Behavior, 52*(2), 424–459.

Bhatt, M. A., Lohrenz, T., Camerer, C. F., & Montague, P. R. (2012). Distinct contributions of the amygdala and parahippocampal gyrus to suspicion in a repeated bargaining game. *PNAS Proceedings of the National Academy of Sciences of the United States of America, 109,* 8728–8733.

Bechara, A., & Damasio, A. R. (2005). The somatic marker hypothesis: A neural theory of economic decision. *Games and Economic Behavior, 52,* 336–372.

Beugré, C. D. (2009). Exploring the neural basis of fairness: A model of neuro-organizational justice. *Organizational Behavior and Human Decision Processes, 110*(2), 129–139.

Binmore, K. (2007). *Does game theory work? The bargaining challenge*. Cambridge, MA: MIT Press.

Bohnet, I., & Zeckhauser, R. (2004). Trust, risk and betrayal. *Journal of Economic Behavior & Organization, 55*(4), 467–484.

Breiter, H. C., Aharon, I., Kahneman, D., Dale, A., & Shizgal, P. (2001). Functional imaging of neural responses to expectancy and experience of monetary gains and losses. *Neuron, 30,* 619–639.

Buehler, R., McFarland, C., Spyropoulos, V., & Lam, K. C. (2007). Motivated prediction of future feelings: effects of negative mood and mood orientation on affective forecasts. *Personality and Social Psychology Bulletin, 33*(9), 1265–1278.

Cabeza, R. & Nyberg, L. (2000). Imaging cognition II: An empirical review of 275 PET and FMRI studies. *Journal of Cognitive Neuroscience, 12*, 1–47.

Cador, M., Robbins, T., & Everitt, B. (1989). Involvement of the amygdala in stimulus reward associations: interaction with the ventral striatum. *Neuroscience, 30*(1), 77–86.

Camerer, C. F. (2013). Goals, Methods, and Progress in Neuroeconomics. *Annual Review of Economics, 5,* 425–455.

Camerer, C. F., & Hare, T. A. (2014). The neural basis of strategic choice. In P. W. Glimcher & E. Fehr (Eds.), *Neuroeconomics: Decision making and the brain* (2nd ed., pp. 479–492). New York: Academic Press.

Caplin, A., & Leahy, J. (2001). Psychological expected utility theory and anticipatory feelings. *The Quarterly Journal of Economics, 116*(1), 55–79.

Colander, D. (2007). Retrospectives: Edgeworth's hedonimeter and the quest to measure utility. *Journal of Economic Perspectives, 21*(2), 215–225.

Chark, R., & Chew, S. H. (2015). A neuroimaging study of preference for strategic uncertainty. *Journal of Risk and Uncertainty, 50*(3), 209–227.

Chib, V. S., Rangel, A., Shimojo, S., & O'Doherty, J. P. (2009). Evidence for a common representation of decision values for dissimilar goods in human ventromedial prefrontal cortex. *Journal of Neuroscience, 29,* 12315–12320.

Chib, V. S., De Martino, B., Shimojo, S., & O'Doherty, J. P. (2012). Neural mechanisms underlying paradoxical performance for monetary incentives are driven by loss aversion. *Neuron, 74*(3), 582–594.

Cikara, M., Botvinick, M. M. & Fiske, S. T. Us versus them: Social identity shapes neural responses to intergroup competition and harm. *Psychological Science, 22*, 306–313 (2011).

Coricelli, G., Critchley, H. D., Joffily, M., O'Doherty, J. P., Sirigu, A. & Dolan, R. J. (2005). Regret and its avoidance: A neuroimaging study of choice behavior. *Nature Neuroscience, 8*, 1253–1262.

Coricelli, G., Dolan, R.J. & Sirigu, A. (2007), Brain, emotion and decision making: The paradigmatic example of regret. *Trends in Cognitive Sciences, 11*(6), 258–265.

Crockett, M. J., & Fehr, E. (2014). Pharmacology of economic and social decision making. In P. W. Glimcher & E. Fehr (Eds.), *Neuroeconomics: Decision making and the brain* (2nd ed., pp. 259–279). New York: Academic Press.

Daw, N. D., & O'Doherty, J. P. (2014). Multiple Systems for Value Learning. In P. W. Glimcher & E. Fehr (Eds.), *Neuroeconomics: Decision making and the brain* (2nd ed., pp. 393–410). New York: Academic Press.

Daw, N. D., & Tobler, P. N. (2014). Value learning through reinforcement: the basics of dopamine and reinforcement learning. In P. W. Glimcher & E. Fehr (Eds.), *Neuroeconomics: Decision making and the brain* (2nd ed., pp. 283–298). New York: Academic Press.

Daw, N. D., Gershman, S. J., Seymour, B., Dayan, P., & Dolan, R. J. (2011). Model-based influences on humans' choices and striatal prediction errors. *Neuron, 69*(6), 1204–1215.

Dayan, P., & Niv, Y. (2008). Reinforcement learning: The good, the bad and the ugly. *Current Opinion in Neurobiology, 18,* 1–12.

Dayan, P., & Seymour, B. (2009). Values and actions in aversion. In P. W. Glimcher, C. F. Camerer, E. Fehr, & R. A. Poldrack (Eds.), *Neuroeconomics: Decision making and the brain*. New York: Academic Press.

Delgado, M.R., Miller, M.M., Inati, S., & Phelps, E.A. (2005). An fMRI study of reward- related probability learning. *NeuroImage, 24*(3), 862–873.

De Martino, B., Camerer, C., & Adolphs, R. (2010). Amygdala damage eliminates monetary loss aversion. *PNAS Proceedings of the National Academy of Sciences of the United States of America, 107*(8), 3788–3792.

De Martino, B., Kumaran, D., Seymour, B., & Dolan, R. J. (2006). Frames, biases, and rational decision-making in the human brain. *Science, 313,* 684–687.

Denk, F., Walton, M. E., Jennings, K. A., Sharp, T., Rushworth, M. F., & Bannerman, D. M. (2005). Differential involvement of serotonin and dopamine systems in cost-benefit decisions about delay or effort. *Psychopharmacology (Berl), 179,* 587–596.

DeQuervain, D., Fischbacher, U., Treyer, V., Schellhammer, M., Schnyder, U., Buck, A., et al. (2004). The neural basis of altruistic punishment. *Science, 305,* 1254–1258.

Ding, L., & Hikosaka, O. (2006). Comparison of reward modulation in the frontal eye field and caudate of the macaque. *Journal of Neuroscience, 26,* 6695–6703.

Dorris, M. C., & Glimcher, P. W. (2004). Activity in posterior parietal cortex is correlated with the subjective desirability of an action. *Neuron, 44,* 365–378.

Dvash, J., Gilam, G., Ben-Ze'ev, A., Hendler, T. & Shamay-Tsoory, S. G. (2010). The envious brain: The neural basis of social comparison. *Human Brain Mapping, 31*, 1741–1750.

de Wit, S., Watson, P., Harsay, H. A., Cohen, M. X., van de Vijver, I., & Ridderinkhof, K. R. (2012). Corticostriatal connectivity underlies individual differences in the balance between habitual and goal-directed action control. *Journal of Neuroscience, 32,* 12066–12075.

Efremidze, L., Sarraf, G., Miotto, K., & Zak, P. J. (2017). The neural inhibition of learning increases asset market bubbles: Experimental evidence. *Journal of Behavioral Finance, 18*(1), 114–124.

Egidi, G., Nusbaum, H. C., & Cacioppo, J. T. (2008). Neuroeconomics: Foundational issues and consumer relevance. In C. P. Haugtvedt, P. M. Herr, & F. R. Kardes (Eds.), *Handbook of Consumer Psychology* (pp. 1177–1214). New York: Psychology Press.

Eisenegger, C., Knoch, D., Ebstein, R. P., Gianotti, L. R., Sa´ndor, P. S., & Fehr, E. (2010). Dopamine receptor D4 polymorphism predicts the effect of L-DOPA on gambling behavior. *Biological Psychiatry, 67*(8), 702–706.

Ekins, W. G., Caceda, R., Capra, C. M., & Berns, G. S. (2013). You cannot gamble on others: Dissociable systems for strategic uncertainty and risk in the brain. *Journal of Economic Behavior & Organization, 94,* 222–233.

Everitt, B. J., & Robbins, T. W. (2005). Neural systems of reinforcement for drug addiction: from actions to habits to compulsion. *Nature Neuroscience, 8*(11), 1481–1489.

Fecteau, S., Knoch, D., Fregni, F., Sultani, N., Boggio, P., & Pascual-Leone, A. (2007). Diminishing risk-taking behavior by modulating activity in the prefrontal cortex: A direct current stimulation study. *Journal of Neuroscience, 27,* 12500–12505.

Fehr, E., & Camerer, C. F. (2007). Social neuroeconomics: The neural circuitry of social preferences. *Trends in Cognitive Science, 11,* 419–427.

Fehr, E., Fischbacher, U., & Kosfeld, M. (2005). Neuroeconomic foundations of trust and social preferences: Initial evidence. *Neuroscientific Foundations of Economic Decision-Making, 95*(2), 346–351.

Fehr, E. & Gächter, S. (2000). Fairness and retaliation: The economics of reciprocity. *Journal of Economic Perspectives, 14,* 159–181.

Fodor, J. A. (1983). *The Modularity of Mind*. Cambridge, MA: MIT Press.

Forgas, J. (1995). Mood and judgment: The affect infusion model (AIM). *Psychology Bulletin, 117,* 39–66.

Fox, C. R., & Poldrack, R. A. (2009). Prospect theory and the brain. In P. W. Glimcher, C. F. Camerer, E. Fehr, & R. A. Poldrack (Eds.), *Neuroeconomics: Decision making and the brain*. New York: Academic Press.

Frydman, C., Barberis, N., Camerer, C. F., Bossaerts, P., & Rangel, A. (2014). Testing theories of investor behavior using neural data. *Journal of Finance, 69*(2), 907–946.

Fumagalli, R. (2014). Neural findings and economic models: Why brains have limited relevance for economics. *Philosophy of the Social Sciences, 44*(5), 606–629.

Gigerenzer, G., Todd, P. M., & ABC Research Group. (1999). *Simple heuristics that make us smart*. New York: Oxford University Press Inc.

Glimcher, P. W. (2003). *Decisions, uncertainty, and the brain: The science of neuroeconomics*. Cambridge, MA: MIT Press.

Glimcher, P. W. (2011). *Foundations of Neuroeconomic Analysis.* New York, New York: Oxford University Press.

Glimcher, P. W. (2014a). Introduction to neuroscience. In P. W. Glimcher & E. Fehr (Eds.), *Neuroeconomics: Decision making and the brain* (2nd ed., pp. 63–75). New York: Academic Press.

Glimcher, P. W. (2014b). Value-based decision making. In P. W. Glimcher & E. Fehr (Eds.), *Neuroeconomics: Decision making and the brain* (2nd ed., pp. 373–391). New York: Academic Press.

Glimcher, P. W. & Fehr, E. (2014). Introduction: A brief history of neuroeconomics. In P. W. Glimcher, & E. Fehr, (Eds.), *Neuroeconomics: Decision making and the brain*, (2nd ed., pp. xvii–xxviii). New York: Academic Press.

Glimcher, P. W., & Rustichini, A. (2004). Neuroeconomics: The consilience of brain and decision. *Science, 306,* 447–452.

Goetz, J., & James, R. N., III. (2008). Human choice and the emerging field of neuroeconomics: A review of brain science for the financial planner. *Journal of Personal Finance, 6*(4), 13–36.

Gospic, K., Mohlin, E., Fransson, P., Petrovic, P., Johanesson, M., & Ingvar, M. (2011). Limbic justice amygdala involvement in immediate rejection in the ultimatum game. *PLoS Biology, 9*(5), 1–8.

Gottfried, J. A., O'Doherty, J., & Dolan, R. J. (2003). Encoding predictive reward value in human amygdala and orbitofrontal cortex. *Science, 301*(5636), 1104–1107.

Hare, T. A., Camerer, C. F., & Rangel, A. (2009). Self-control in decision-making involves modulation of the vmPFC valuation system. *Science, 324*(5927), 646–648.

Hariri, A. R., Mattay, V. S., Tessitore, A., Fera, F., Smith, W. G., & Weinberger, D. R. (2002). Dextroamphetamine modulates the response of the human amygdala. *Neuropsychopharmacology, 27*(6), 1036–1040.

Harrison, G. W. (2008). Neuroeconomics: A critical reconsideration. *Economics and Philosophy, 24,* 303–447.

Hochman, G., & Yechiam, E. (2011). Loss aversion in the eye and in the heart: The autonomic nervous system's responses to losses. *Journal of Behavioral Decision Making, 24*(2), 140–156.

Hodgkinson, G. P., & Healey, M. P. (2011). Psychological foundations of dynamic capabilities: reflexion and reflection in strategic management. *Strategic Management Journal, 32*(13), 1500–1516.

Houser, D., & McCabe, K. (2014). Experimental economics and experimental game theory. In P. W. Glimcher & E. Fehr (Eds.), *Neuroeconomics: Decision making and the brain* (2nd ed., pp. 19–34). New York: Academic Press.

Hsu, M., Bhatt, M., Adolphs, R., Tranel, D., & Camerer, C. F. (2005). Neural systems responding to degrees of uncertainty in human decision-making. *Science, 310,* 1680–1683.

Huettel, S. A. & Payne, J. W. (2009). Commentaries and rejoinder to "trade-off aversion as an explanation for the attraction effect: A functional magnetic resonance imaging study. *Journal of Marketing Research, XLVI,* 14–24.

Huettel, S. A., Stowe, C. J., Gordon, E. M., Warner, B. T., & Platt, M. L. (2006). Neural signatures of economic preferences for risk and ambiguity. *Neuron, 49,* 765–775.

Hutzler, F. (2014). Reverse inference is not a fallacy per se: Cognitive processes can be inferred from functional imaging data. *NeuroImage, 84,* 1061–1069.

Kable, J. W. (2011). The cognitive neuroscience toolkit for the neuroeconomist: A functional overview. *Journal of Neuroscience, Psychology, and Economics, 4*, 63–84.

Kable, J. W., & Glimcher, P. W. (2007). The neural correlates of subjective value during intertemporal choice. *Nature Neuroscience, 10*(12), 1625–1633.

Kable, J. W., & Glimcher, P. W. (2010). An "as soon as possible" effect in human inter-temporal decision making: Behavioral evidence and neural mechanisms. *Journal of Neurophysiology, 103,* 2513–2531.

Kahneman, D., Knetsch, J. L., & Thaler, R. H. (1990). Experimental tests of the endowment effect and the Coase theorem. *Journal of Political Economy, 98*(6), 1325–1348.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica, 47,* 263–291.

Knoch, D., Gianotti, L. R., Pascual-Leone, A., Treyer, V., Regard, M., Hohmann, M., et al. (2006a). Disruption of right prefrontal cortex by low-frequency repetitive transcranial magnetic stimulation induces risk-taking behavior. Journal of Neuroscience, 26, 6469–6472.

Knoch, D., Pascual-Leone, A., Meyer, K., Treyer, V., & Fehr, E. (2006b). Diminishing reciprocal fairness by disrupting the right prefrontal cortex. *Science, 314,* 829–832.

Knutson, B., & Greer, S. M. (2008). Anticipatory affect: Neural correlates and consequences for choice. *Philosophical Transactions of the Royal Society of London. Series B, Biological sciences, 363*(1511), 3771–3786.

Knutson, B., Fong, G. W., Bennett, S. M., Adams, C. M., & Hommer, D. (2003). A region of mesial prefrontal cortex tracks monetarily rewarding outcomes: characterization with rapid event-related fMRI. *NeuroImage, 18,* 263–272.

Knutson, B., Wimmer, G., Kuhnen, C., & Winkielman, P. (2008). Nucleus accumbens activation mediates the influence of reward cues on financial risk taking. *NeuroReport, 19*(5), 509–513.

Kosfeld, M., Heinrichs, M., Zak, P., Fischbacher, U., & Fehr, E. (2005). Oxytocin increases trust in humans. *Nature, 435*(7042), 673–676.

Krueger, F., McCabe, K., Moll, J., Kriegeskorte, N., Zahn, R., Strenziok, M., et al. (2007). Neural correlates of trust. *PNAS Proceedings of the National Academy of Sciences of the United States of America, 104*(50), 20084–20089.

Kuhnen, C. M., & Knutson, B. (2011). The influence of affect on beliefs, preferences, and financial decisions. *Journal of Financial and Quantitative Analysis, 46,* 605–626.

Lambert, B., Declerck, C. H., Boone, C., Parizel, P. M. (2017). How oxytocin fine-tunes decision making in social dilemmas: cooperate as long as it pays off, but aggress only when you think you can win! [Abstract]. In *Proceedings from 2016 NeuroPsychoEconomics* Conference. Bonn: Association for NeuroPsychoEconomics.

Lau, B., & Glimcher, P. W. (2008). Value representations in the primate striatum during matching behavior. *Neuron, 58,* 451–463.

Lempert, K. M., & Phelps, E. A. (2014). Neuroeconomics of emotion and decision making. In P. W. Glimcher & E. Fehr (Eds.), *Neuroeconomics: Decision making and the brain* (2nd ed., pp. 219–236). New York: Academic Press.

Levitt, S. D., & List., J. A. (2007). What do laboratory experiments measuring social preferences reveal about the real world? *Journal of Economic Perspectives, 21*, 153–174.

Levy, D. J., & Glimcher, P. W. (2011). Comparing apples and oranges: Using reward-specific and reward-general subjective value representation in the brain. *The Journal of Neuroscience, 31*(41), 14693–14707.

Levy, D. J., & Glimcher, P. W. (2012). The root of all value: A neural common currency for choice. *Current Opinion in Neurobiology, 22*(6), 1027–1038.

Levy, I., Snell, J., Nelson, A. J., Rustichini, A., & Glimcher, P. W. (2010). Neural representation of subjective value under risk and ambiguity. *Journal of Neurophysiology, 103*(2), 1036–1047.

Lin, C.-H., Tuan, H.-P., & Chiu, Y.-C. (2010). Medial frontal activity in brand-loyal consumers: A behavior and near-infrared ray study. *The Journal of Neuroscience, Psychology, and Economics, 3,* 59–73.

Lo, A. W., & Repin, D. V. (2002). The psychophysiology of real-time financial risk processing. *Journal of Cognitive Neuroscience*, 14, 323–339.

Lori, R. (2017). Sex differences through a neuroscience lens: Implications for business ethics. *Journal of Business Ethics, 144*(4), 771–782.

Louie, K., & De Martino, K. (2014). The neurobiology of context-dependent valuation and choice. In P. W. Glimcher & E. Fehr (Eds.), *Neuroeconomics: Decision making and the brain* (2nd ed., pp. 455–476). New York: Academic Press.

Louie, K., & Glimcher, P. W. (2010). Separating value from choice: Delay discounting activity in the lateral intraparietal area. *Journal of Neuroscience, 30*(16), 5498–5507.

Masten, C. L., Eisenberger, N. I., Borofsky, L. A., Pfeifer, J. H., McNealy, K., Mazziotta, J. C., et al. (2009). Neural correlates of social exclusion during adolescence: Understanding the distress of peer rejection. *Social Cognitive and Affective Neuroscience, 4*(2), 143–157.

McCabe, K., & Singer, T. (2008). Brain signatures of social decision-making. In: C. Engel, Singer, W. (Eds.), *Better than conscious: Exploiting the capacity of humans to reach decisions by both serial and parallel processing of information* (pp. 103–123). Cambridge, MA: MIT Press.

McFadden, D. L. (2005). Revealed stochastic preference: A synthesis. *Economic Theory, 26*(2), 245–264.

Menz, M. M., Büchel, C., & Peters, J. (2012). Sleep deprivation is associated with attenuated parametric valuation and control signals in the midbrain during value-based decision making. *Journal of Neuroscience, 32,* 6937–6946.

Miu, A. C., & Crişan, L. G. (2011). Cognitive reappraisal reduces the susceptibility to the framing effect in economic decision making. *Personality and Individual Differences, 51,* 478–482.

Mohr, P. N., Biele, G., & Heekeren, H. R. (2010). Neural processing of risk. *Journal of Neuroscience, 30,* 6613–6619.

Nakahara, H., Itoh, H., Kawagoe, R., Takikawa, Y., & Hikosaka, O. (2004). Dopamine neurons can represent context-dependent prediction error. *Neuron, 41,* 269–280.

Newsome, W. T., Britten, K. H., & Movshon, J. A. (1989). Neuronal correlates of a perceptual decision. *Nature, 341,* 52–54.

Nieuwenhuis, S., Heslenfeld, D. J., von Geusau, N. J., Mars, R. B., Holroyd, C. B., & Yeung, N. (2005). Activity in human reward-sensitive brain areas is strongly context dependent. *NeuroImage, 25*(4), 1302–1309.

Niv, Y., Edlund, J. A., Dayan, P., & O'Doherty, J. P. (2012). Neural prediction errors reveal a risk-sensitive reinforcement-learning process in the human brain. *Journal of Neuroscience, 32*(2), 551–562.

Noonan, M. P., Walton, M. E., Behrens, T. E., Sallet, J., Buckley, M. J., & Rushworth, M. F. (2010). Separate value comparison and learning mechanisms in macaque medial and lateral orbitofrontal cortex. *PNAS Proceedings of the National Academy of Sciences of the United States of America, 107*(47), 20547–20552.

O'Doherty, J., Deichmann, R., Critchley, H. D., & Dolan, R. J. (2002). Neural responses during anticipation of a primary taste reward. *Neuron, 33*(5), 815–826.

Ongur, D., & Price, J. L. (2000). The organization of networks within the orbital and medial prefrontal cortex of rats, monkeys and humans. *Cerebral Cortex, 10*(3), 206–219.

Ortmann, A. (2008). Prospecting neuroeconomics. *Economics and Philosophy, 24,* 431–448.

Paulus, M. P., Feinstein, J. S., Castillo, G., Simmons, A. N., & Stein, M. B. (2005). Dose-dependent decrease of activation in bilateral amygdala and insula by lorazepam during emotion processing. *Archives of General Psychiatry, 62*(3), 282 – 288.

Pavlov, I. P. (1927). *Conditioned Reflexes: An Investigation of the Physiological Activity of the Cerebral Cortex.* (G. V. Anrep., Trans.) London: Oxford University Press.

Pessiglione, M., Seymour, B., Flandin, G., Dolan, R. J., & Firth, C. D. (2006). Dopamine-dependent prediction errors underpin reward-seeking behaviour in humans. *Nature, 442*(7106), 1042–1045.

Peters, J., & Büchel, C. (2010). Episodic future thinking reduces reward delay discounting through an enhancement of prefrontal-mediotemporal interactions. *Neuron, 66*, 138–148.

Pine, A., Seymour, B., Roiser, J. P., Bossaerts, P., Friston, K. J., Curran, H. V., et al. (2009). Encoding of marginal utility across time in the human brain. *Journal of Neuroscience, 29*(30), 9575–9581.

Plassmann, H., O'Doherty, J., Shiv, B., & Rangel, A. (2008). Marketing actions can modulate neural representations of experienced pleasantness. *PNAS Proceedings of the National Academy of Sciences of the United States of America, 105*(3), 1050–1054.

Platt, M. L., & Glimcher, P. W. (1999). Neural correlates of decision variables in parietal cortex. *Nature, 400,* 233–238.

Platt, M. L., & Plassmann, H. (2014). Multistage valuation signals and common neural currencies. In P. W. Glimcher & E. Fehr (Eds.), *Neuroeconomics: Decision making and the brain* (2nd ed., pp. 237–258). New York: Academic Press.

Porcelli, A. J., & Delgado, M. R. (2009). Acute stress modulates risk taking in financial decision making. *Psychological Science, 20,* 278–283.

Poldrack, R. (2006). Can cognitive processes be inferred from neuroimaging data? *Trends in Cognitive Sciences, 10*(2), 59–63.

Poldrack, R. A. (2011). Inferring mental states from neuroimaging data: from reverse inference to large-scale decoding. *Neuron, 72*(5), 692–697.

Powell, T. C. (2011). Neurostrategy. *Strategic Management Journal, 32*, 1484–1499.

Prevost, C., Pessiglione, M., Metereau, E., Clery-Melin, M. L., & Dreher, J. C. (2010). Separate valuation subsystems for delay and effort decision costs. *Journal of Neuroscience, 30*(42), 14080–14090.

Prevost, C., Liljeholm, M., Tyszka, J. M., & O'Doherty, J. P. (2012). Neural correlates of specific and general Pavlovian-to-Instrumental Transfer within human amygdala subregions: A high-resolution fMRI study. *Journal of Neuroscience, 32*(24), 8383–8390.

Racine, E., Bar-Ilan, O., & Illes, J. (2005). fMRI in the public eye. *Nature Reviews Neuroscience, 6*(2), 159–164.

Rangel, A., & Hare, T. (2010). Neural computations associated with goal-directed choice. *Current Opinion in Neurobiology, 20*(2), 262–270.

Read, D., & Read, N. L. (2004). Time discounting over the life- span. *Organizational Behavior and Human Decision Processes, 94,* 22–32.

Rilling, J. K., Gutman, D. A., Zeh, T. R., Pagnoni, G., Berns, G. S., & Kilts, C. D. (2002). A neural basis for social cooperation. *Neuron, 35*(2), 395–405.

Rubinstein, A. (2006). Discussion of 'Behavioral Economics'. In R. Blundell, W. K. Newey, & T. Persson (Eds.), *Advances in Economics and Econometric Theory (vol II)* (pp. 246–254). New York: Cambridge University Press.

Ruff, C. C., & Huettel, S. A. (2014). Experimental methods in cognitive neuroscience. In P. W. Glimcher & E. Fehr (Eds.), *Neuroeconomics: Decision making and the brain* (2nd ed., pp. 77–108). New York: Academic Press.

Samuelson, P. A. (1938). A note on the pure theory of consumer behavior. *Economica, 1,* 61–71.

Sanfey, A., Rilling, J., Aronson, J., Nystrom, L., & Cohen, J. (2003). The neural basis of economic decision-making in the Ultimatum game. *Science, 300,* 1755–1758.

Schwarz, N., & Clore, G. (1983). Mood, misattribution, and judgments of well-being: Informative and directive functions of affective states. *Journal of Personality and Social Psychology, 45*(3), 513–523.

Simon, H. A. (1955). A behavioral model of rational choice. *The Quarterly Journal of Economics, 69,* 99–118.

Simon, H. A. (1997). *An Empirically Based Microeconomics*. Cambridge, UK: Cambridge University Press.

Singer, T., & Fehr, E. (2005). The neuroeconomics of mind reading and empathy. *Neuroscientific Foundations of Economic Decision-Making, 95*(2), 340–345.

Singer, T., & Tusche, A. (2014). Understanding others: brain mechanisms of theory of mind and empathy. In P. W. Glimcher & E. Fehr (Eds.), *Neuroeconomics: Decision making and the brain* (2nd ed., pp. 513–532). New York: Academic Press.

Smith, K., Dickhaut, J., McCabe, K., & Pardo, J. V. (2002). Neuronal substrates for choice under ambiguity, risk, gains, and losses. *Management Science, 48*(6), 711–718.

Soman, D., Ainslie, G., Frederick, S., Li, X., Lynch, J., Moreau, P., et al. (2005). The psychology of intertemporal discounting: Why are distant events valued differently from proximal ones? *Marketing Letters,* (3/4), 347–360.

Sokol-Hessner, P., Hsu, M., Curley, N. G., Delgado, M. R., Camerer, C. F., & Phelps, E. A. (2009). Thinking like a trader selectively reduces individuals' loss aversion. *PNAS Proceedings of the National Academy of Sciences of the United States of America, 106,* 5035–5040.

Sokol-Hessner, P., Camerer, C. F., & Phelps, C. A. (2013). Emotion regulation reduces loss aversion and decreases amygdala responses to losses. *Social Cognitive and Affective Neuroscience, 8*(3), 341–350.

Steinbeis, N., Bernhardt, B., & Singer, T. (2012). Impulse control and underlying functions of the left DLPFC mediate age-related and age-independent individual differences in strategic social behavior. *Neuron, 73,* 1040–1051.

Steiner, A. P., & Redish, A. D. (2014). Behavioral and neurophysiological correlates of regret in rat decision-making on a neuroeconomic task. *Nature Neuroscience, 17*(7), 995–1004.

Talmi, D., Seymour, B., Dayan, P., & Dolan, R. J. (2008). Human Pavlovian-instrumental transfer. *Journal of Neuroscience, 28*(2), 360–368.

Takahashi, T., Hadzibeganovic, T., Cannas, S. A., Makino, T., Hiroki Fukui, H., & Kitayama, S. (2010). Cultural Neuroeconomics of Intertemporal Choice. *Journal of Behavioral Economics and Finance, 3,* 133–135.

Tanaka, S. C., Doya, K., Okada, G., Ueda, K., Okamoto, Y., & Yamawaki, S. (2004). Prediction of immediate and future rewards differently recruits cortico-basal ganglia loops. *Nature Neuroscience, 7*(8), 887–893.

Thaler, R. H., & Shefrin, H. M. (1981). An economic theory of self-control. *Journal of Political Economy, 89*(2), 392–406.

Tobler, P. N., & Weber, E. U. (2014). Valuation for risky and uncertain choices. In P. W. Glimcher & E. Fehr (Eds.), *Neuroeconomics: Decision making and the brain* (2nd ed., pp. 149–171). New York: Academic Press.

Tom, S. M., Craig, R. F., Trepel, C., & Poldrack, R. A. (2007). The neural basis of loss aversion in decision-making under risk. *Science, 315*(5811), 515–518.

Tricomi, E., Balleine, B. W., & O'Doherty, J. P. (2009). A specific role for posterior dorsolateral striatum in human habit learning. *European Journal of Neuroscience, 29,* 2225–2232.

Tversky, A., & Kahneman, D. (1992). Advances in prospect-theory cumulative representation of uncertainty. *Journal of Risk and Uncertainty, 5*(4), 297–323.

Veblen, T. (1898). Why is economics not an evolutionary science? *Quarterly Journal of Economics, 12,* 373–397.

Vercoe, M., & Zak, P. J. (2010). Inductive modeling using causal studies in neuroeconomics: Brains on drugs. *Journal of Economic Methodology, 17*(2), 133–146.

Volk, S., & Köhler, T. (2012). Brains and games: Applying neuroeconomics to organizational research. *Organizational Research Methods, 15*(4), 522–552.

Votinov, M., Pripfl, J., Windischberger, C., Sailer, U., & Lamm, C. (2015). Better you lose than I do: Neural networks involved in winning and losing in a real time strictly competitive game. *Scientific Reports, 5,* 1–13.

Wallis, J. D., & Rushworth, M. F. S. (2014). Integrating benefits and costs in decision making. In P. W. Glimcher & E. Fehr (Eds.), *Neuroeconomics: Decision making and the brain* (2nd ed., pp. 411–433). New York: Academic Press.

Weber, B., Aholt, A., Neuhaus, C., Trautner, P., Elger, C. E., & Teichert, T. (2007). Neural evidence for reference-dependence in real-market-transactions. *NeuroImage, 35*(1), 441 – 447.

Witt, U., & Binder, M. (2013). Disentangling motivational and experiential aspects of "utility"- A neuroeconomics perspective. *Journal of Economic Psychology, 36,* 27–40.

Yacubian, J., Gläscher, J., Schroeder, K., Sommer, T., Braus, D.F., & Bu¨chel, C. (2006). Dissociable systems for gain-and loss-related value predictions and errors of prediction in the human brain. *Journal of Neurocience, 26,* 9530–9537.

Zhao, Y., & Siau, K. (2016). Cognitive Neuroscience in Information Systems Research. *Journal of Database Management, 27*(1), 58–73.

Information and Communication Technology: Understanding Their Dark-Side Effects

Monideepa Tarafdar and Jean-François Stich

"We are letting technology take us places that we don't want to go"

Sherry Turkle, TED talk, 2012, https://youtu.be/t7Xr3AsBEK4?t=144

1 The Dark Side of Information and Communication Technologies

Information and Communication Technologies, hereafter referred to as ICT, are a double-edged sword. While they have been used to generate unquestionable benefits for organizations and societies (e.g. efficiency in information processing, innovations in business processes, better products and services[1]), we are now seeing a plethora of negative phenomenon associated with them. Increasingly referred to as the dark side of ICT use (e.g. Tarafdar et al. 2015a), these include among others, technostress, technology addiction, and information and communication overload. As these types of phenomena emerge and become observable, they are generating a buzz of

[1] The benefits have been analysed and explained at length for the past four to five decades in various management disciplines, most notably in Management Information Systems, Organizational Behaviour, Strategic Management and Operations Management. The interested reader is directed to the literature (i.e. journals, conferences and texts) in these disciplines.

M. Tarafdar (✉)
Lancaster University (Management School), Lancaster, UK
e-mail: m.tarafdar@lancaster.ac.uk

J.-F. Stich
ICN Business School, Nancy, France
e-mail: jfstich@icn-groupe.fr

P. Ray et al. (eds.), *Economics, Management and Sustainability*,
https://doi.org/10.1007/978-981-13-1894-8_14

substantive and considerable scholarly commentary. For example, we know that '*In a society that is constantly speeding up, technology has provided excellent means for achieving competition-driven goals. But in so far as it affects employees, it may have given rise to unexpected side effects such as stress and distress*' (Mano and Mesch 2010, p. 67). What is ironic is that the very qualities of ICT that make them useful, such as constant availability and reliability, quick and complex information processing capabilities, and user-friendly interfaces, are also key enablers of these dark-side phenomena. For example, while we value the ease and convenience of on-demand email, we also know that it is costly in terms of time and productivity. For instance, the daily cost of reading email for a large UK firm with 2850 email users is £40,848, and the cost per year over £9.8 million (Jackson et al. 2006).

Shoshana Zuboff, in 1988, described '*informating*' (Zuboff 1988), the ability to track data as they are generated and processed, as a double-edged effect that can both lead to efficient work processes and isolation and control of individuals who execute them. As early as in the 1960s, Marshall McLuhan warned us that the '*medium is the message*', in that although we might start off by designing and shaping our information processing and communication tools according to our needs and requirements, eventually they would shape what we do, in unexpected and unforeseen ways. Winding through the technology maze to the present, in 2018, the effects of these phenomena are beginning to be seen among individuals as well as collectives, spanning organizations and business enterprises, and society at large. We examine in this chapter, what some of these dark-side phenomena are, how they affect our social and business enterprises, and what can be done by way of their mitigation.

2 Dark-Side Phenomena: Technostress, Technology Addiction, Overload from Computer-Mediated Communication

Technostress is the stress that individuals experience due to their use of ICT (e.g. Tarafdar et al. 2017). This phenomenon illustrates that ICT place various demands on the individual (e.g. Cooper et al. 2001; McGrath 1976), which the individual appraises as threats that he or she is unable to meet, and faces the possibility of adverse consequences as a result. The very characteristics of ICT that we value such as their ubiquity, reliability, ease of use, mobility and constant presence can present such demands, as can ICT-related events such as system breakdown and ICT-created interruptions (Galluch et al. 2015). When faced with these demands, the individual experiences the presence of technostress creating conditions, known as techno-stressors. The literature reveals a number of techno-stressors (e.g. Ragu-Nathan et al. 2008; Tarafdar et al. 2007). Techno-overload forces the user to do more in order to use the technology, such as understanding and using a plethora of features, and having to adhere to extra requirements such as security-related tasks (D'Arcy et al. 2014; Reinke et al. 2014; Tarafdar et al. 2007). Techno-invasion is

the stressor because of which individuals feel that their non-work time is invaded by work demands (Tarafdar et al. 2007); they are faced with expectations of constant availability and immediate response vis-à-vis their work (Barber and Santuzzi 2015; Day et al. 2012). Individuals experience techno-uncertainty when they feel that the ICT they use change quickly and they have little control over the change-related decisions, use and policies (Tarafdar et al. 2007). Techno-insecurity embodies the feeling of insecurity that individuals face when they feel that their colleagues may know more about new technologies than they do (Tarafdar et al. 2007). Techno-complexity is the techno-stressor that individuals face when they find ICT difficult to use and find themselves on a constant learning curve in order to use frequently changing ICT, understand their complications and deal with the hassles in dealing with them (Barber and Santuzzi 2015; Barley et al. 2011, p. 201; Day et al. 2012; Galluch et al. 2015; Tarafdar et al. 2007). Individuals who are digitally literate and involved, and have a positive attitude towards ICT to begin with, may experience lower levels of techno-stressors (Barley et al. 2011; Tarafdar et al. 2015b; Tarafdar et al. 2010). One application that is known to be an important cause of technostress is social media such as Facebook. This is because they come with expectations to attend to social needs of others, and to produce and consume excessive information about other people's lives which can be difficult to process. At the same time, not doing so can lead to an experience of fear of missing out on interesting updates if one cannot process such information and an overload of social media-related activities such as posting and browsing at the cost of others (Maier et al. 2014).

Technology addiction happens when individuals engage in pathological use of ICT. Such use stems from a maladaptive and harmful dependency on ICT. It tends to be obsessive compulsive and out of control, and to the detriment of other important activities and commitments (Turel et al. 2011; Xu et al. 2012). Addiction to ICT often embodies a dangerous and self-fulfilling promise—those with higher levels of addiction to a particular application consider that application to be more useful, easier and more fun to use, than non-addicts (Turel and Serenko 2012)—which has the potential to lead to never-ending cycle of addictive use of ICT. The American Psychiatric Association describes it as 'pathological computer use' (Young 2010, p. 91). How addicted an individual is may depend on their social-demographics, personality traits (Kim et al. 2008), whether or not the ICT applications they use provides opportunities for hedonic and entertainment-related use and to what extent it satisfies the individual's needs (Young 2010). Individuals may get addicted to the use of ICT, especially when they enjoy using it. What is more, as is the case with other forms of addiction, they may be subject to withdrawal and personal relapse if they stop using those applications. As in the case of technostress, the most common ICT that people are addicted to are social networking applications such as Facebook (Turel and Serenko 2012; Xu et al. 2012). Other ICT include the Internet in general, online gaming and, ironically enough, work email (Weinstein and Lejoyeux 2010; Xu et al. 2012). Among ICT devices that individuals are most addicted to are laptops and smartphones, ostensibly because they can run a number of different applications and through that can transport us into different worlds. Indeed, smartphone addiction has emerged as one of the most widespread forms of IT addiction (Montag et al. 2016).

We see behaviours such as employees hooked inexorably to their smartphones, never being disconnected from their work, and browsing work-related email or social media conversations as among the last things they do before going to sleep and the first thing after waking up (Montag et al. 2016).

The dark side of ICT is also revealed in an overload of computer-mediated communication such as use of email. It has received considerable newspaper coverage in recent years, with titles such as 'checking your emails outside of work really IS bad for your health' (capital letters in original) (Davies 2015). Computer-mediation communication is associated with a wide range of detrimental outcomes (Stich et al. 2015). Because of its pervasiveness, employees are increasingly 'working anytime, anywhere' (Eurofound and the International Labour Office 2017) and are expected to remain constantly available, potentially resulting in increased work life imbalance (Derks et al. 2015), job burnout and health-related problems (Barber and Santuzzi 2015). Many choose of their own will to become and remain available through ICT 'around-the-clock' (Matusik and Mickel 2011), seeing it as an imperative for professional advancement. Some, referred to as 'Crackberries', even wait for *the* awaited communication (Mazmanian et al. 2005) and constantly check their smartphones as they do so. ICT applications can be used quickly and conveniently; individuals thus also face ever-increasing volumes of communication, and, therefore, increased feelings of overload (Dabbish and Kraut 2006) and more frequent interruptions (Jackson et al. 2006). However, and at the same time, such applications are not very suitable for efficiently transmitting the rich panel of human emotions and visual cues. As a result, cyber rudeness abounds and yet at the same time civil messages can be misinterpreted, leading to even higher levels of frustration, stress and conflict escalations (Friedman and Currall 2003). Although very few would dispute that computer-mediated communication has facilitated dispersed collaboration and information exchange, the literature surely shows that, insofar that it affects employees, detrimental side effects have also flourished.

3 Effects of Dark-Side Phenomena

These dark-side phenomena are associated with various negative outcomes. Technostress, for example, is associated with a lack of job satisfaction and organizational commitment, intentions to leave one's organization, role overload, role conflict (Ragu-Nathan et al. 2008; Tarafdar et al. 2007), as well as job-related anxiety and depression (Sprigg and Jackson 2006). Further, technostressed individuals are not able to use ICT properly, which is ironical, given that one of the expected outcomes for the implementation of ICT is that people should use them. Such individuals suffer from low abilities to use ICT for innovation and productivity at work, low satisfaction with the ICT they work with (Tarafdar et al. 2010, 2015b) and a lack of willing compliance with requirements such as timely response to email and adherence to security requirements (Barber and Santuzzi 2015; D'Arcy et al. 2014). Even more alarmingly, they suffer from reduced well-being in that they experience increased

exhaustion, burnout and strain (Barber and Santuzzi 2015; Barley et al. 2011; Day et al. 2012; Reinke and Chamorro-Premuzic 2014; Srivastava et al. 2015). Indeed, the physical symptoms are so strong that the incidence of stress hormones such as alpha amylase (Galluch et al. 2015) has been observed when individuals experience ICT-related hassles such as system breakdown. Individuals who experience high levels of techno-stressors may ultimately stop using an ICT application altogether (Maier et al. 2015).

Addiction to ICT similarly comes with a myriad of negative consequences. Perhaps, the most prominent of these is work life conflict (Turel et al. 2011), brought on primarily by addiction to readily available work email on smartphones and tablets. Addiction to mobile phones can also lead to anxiety and insomnia (Jenaro et al. 2007), mood disturbances and antisocial behaviours (Griffiths et al. 2010), together with general psychological distress (Beranuy et al. 2009). In extreme cases, it may lead to lack of satisfaction with the offline (and one would imagine, the real) world (Bruner 2006; Hussain and Griffiths 2009)! Internet addition in general is detrimental to various aspects of the individual's personal life, in the form of depression (Iacovelli and Valenti 2009), loneliness (Morahan-Martin and Schumacher 2000) and reduced self-esteem (Niemz et al. 2005). Within the Internet milieu, another widely prevalent form of ICT addiction is that to social media and instant messaging. In adults, this can cause loss of relationships due to overuse, and in students it has been shown to be associated with low academic performance (Huang and Leung 2009). An extreme and socially dangerous form of addiction is that of cybersex and cyberporn, which can destroy an individual's family life and social relationships, leading to social isolation (Daneback et al. 2005; Schwartz and Southern 2000).

Individuals addicted to ICT can also cause problems for their organizations. Mobile email addiction increases the user's work overload, which in turn reduces his or her organizational commitment and work productivity (Turel and Serenko 2010). Employees addicted to work use of smartphones at night show lower productivity at work the following day (Lanaj et al. 2013). Indeed, employees who become addicted to work-related ICT have been known to litigate against and hold their employers legally responsible, for both work and personal problems stemming from such addiction (Kakabadse et al. 2007)! Overall, therefore, addiction to ICT can reduce the individual's well-being, change his or her social relationships and interaction patterns for the worse, adversely affect his or her work performance, and in doing so create a significant economic burden.

Overload from computer-mediated communication lengthens people's workdays (Barley et al. 2011). Handling email and other communication media is a time-consuming task. Reading email alone can consume a significant portion of the workday (Jackson et al. 2006). Interruptions from computer-mediated communication can lead to significant time loss, given that recovering from such interruptions and re-engaging in the primary task following an interruption can take up to fifteen minutes (Jackson et al. 2001). Such productivity losses are the greatest for employees who highly depend on computer-mediated communication to accomplish their work, given their higher exposal to such (Karr-Wisniewski and Lu 2010). These include employees who telework or work in virtual team working. The social effects of

computer-mediated communication-related overload are even more profound. They include expanding working hours beyond office time effectively 'leashing' employees to their workplaces (Mazmanian et al. 2005). Constant availability enabled by computer-mediated communication is generally associated with increased work life conflict and burnout (Derks et al. 2015; Matusik and Mickel 2011). What is ironical, however, is that in spite of such consequences, individuals might crave for and take pride in constant availability (Matusik and Mickel 2011).

4 Mitigating Dark-Side Phenomena

Fortunately, research also shows that there are ways to mitigate or at least address the negative consequences of these dark-side phenomena. Such mitigation factors can be present at the level of the individual, the organization and the society. However, while a few mitigating mechanisms have been suggested (Tarafdar et al. 2015a, b), research in this area is relatively sparse.

In general, mitigation mechanisms in organizations could be interventions that make a particular dark-side phenomenon less likely to occur such as reducing the number of ICT design features likely to cause stress, overload or addiction for instance. Or, they might be directed at the individual to alleviate his or her dark-side behaviours and attitudes after they are manifested. These could include workplace mechanisms directed towards employees such as education/counselling, awareness, institutional support, job/role redesign and altered reporting structures. In the context of technostress, for example, factors that decrease or inhibit the extent of these negative outcomes include IS management mechanisms such as literacy facilitation, technical support, end user involvement, innovation support, co-worker support and support manuals (Day et al. 2012; Ragu-Nathan et al. 2008; Soucek and Moser 2010). To help employees deal with computer-mediated communication overload, for example, Volkswagen has stopped routing emails to certain employees out of the office (BBC News 2012). A less drastic measure can be to decrease the email checking frequency settings of employees' inboxes in order to reduce exposure to ICT interruptions (Jackson et al. 2006). Some authors go as far as to suggest checking emails only two to four times per day (Gupta et al. 2011). A caveat to such interventions is that emails may pile up in the employees' absence or checking gaps, potentially resulting in further overload (Dabbish and Kraut 2006). A problematic issue, however, is that such one-size-fits-all interventions generally fail to consider that individuals differ in their tolerance and preferences for ICT use. Employees who enjoy and feel empowered remaining constantly available through ICT (Matusik and Mickel 2011) can grow dissatisfied with interventions taking their prided and addictive smartphone away (Mazmanian et al. 2005). However, those employees need to understand that the messages they send and the interactions they initiate may impact their recipients. To accommodate both their preferences and those of others, organizations may favour interventions aiming at raising awareness over ICT-related problems. This can take the form of peer support groups, impactful but time-bound

email bans (Mark et al. 2012), or organizational codes of ICT conduct. Furthermore, email training such as how to write succinct emails has been found to reduce email overload (Soucek and Moser 2010) and may, therefore, deserve to be more developed in organizations.

At the societal level, policies can serve as powerful mitigation mechanisms. For example, we are now seeing government policies potentially shaping IT use activities such as checking email. National policies which are relevant for dark-side IT phenomena may include, for instance, restricting access to potentially harmful or dangerous websites (for instance, child pornography) (Griffiths et al. 2010) and developing appropriate industry-specific regulations, such as in the financial sector. They could also comprise educational and social support at the level of the family and other social units, for creating awareness of IT addiction, cyberbullying, Internet pornography, privacy problems and intellectual property theft.

The personality of the individual may influence the extent to which he or she experiences these dark-side phenomena. For example, technostress has been found to be more prevalent among those who with the personality orientations of neuroticism, agreeableness and extraversion (Srivastava et al. 2015). At the same time, those with greater technology self-efficacy, technology competence, and control over their tasks at work (Tarafdar et al. 2015a, b) experience less technostress. Aside from these, individuals also respond to dark-side phenomenon by actively and constructively coping with them. They may cope through actions that are largely emotional, such as venting, seeking social support and psychological distancing themselves from the ICT and resisting its use (Beaudry and Pinsonneault 2010), in some cases discontinuing the use of stressful ICT altogether (Maier et al. 2015). They may also cope through more instrumental means such as experimenting with new ICT features to fit existing tasks and workflows, consciously sticking to the minimum required use (Beaudry and Pinsonneault 2005), altering their task alterations to suit the requirements of new ICT, adapting their use of ICT (Beaudry and Pinsonneault 2005) and psychologically disengaging themselves from any requirements associated with ICT use through non-compliance (D'Arcy et al. 2014).

5 What Next?

So where do we go from here? Two centuries ago, in what was the dawn of industrialization, textile workers in Northern England raised their own army to wreck their newly introduced machinery. The Luddites rose up fiercely against technology-borne automation of what used to be the blue-collar jobs of the cotton mills. Today, algorithm-based business processes, together with artificial intelligence and cognitive computing, are having similar effects on white-collar knowledge-centric jobs such as law, IT services, financial trading and accounting (e.g. Davenport and Kirby 2016). This comes with unexpected and negative effects. For instance, high-speed financial trading systems have been known to cause 'flash crashes' (Tracy and Patterson 2014). A key reason for these crashes is the use of super-fast, automated trading

algorithms that obtain asymmetrically early access to price-related information published by trading firms through high-speed IT networks. During one such crash that occurred in May 2010, the Dow Jones Index fell an unprecedented 1000 points in the space of a few minutes, as USD 1 trillion was wiped off the value of markets; profits from such crashes often accrue to a miniscule percentage of those who participate in the stock markets.

In parallel, we are witnessing two other developments. One is that the use of ICT now takes place both continually, that is, not confined to a specific physical work or home or leisure time or location, and seamlessly across life activities, that is, not easily classified as 'work' or 'non-work'. The same smartphones, tablets and laptops are used for office email, Facebook browsing, Snapchat posting and WhatsApp texting, in a continual flow of myriad activities that are hard to distinguish or classify. This is particularly more so for digital natives (Vodanovich et al. 2010). And two, over time, individuals get used to the 'upped' doses of ICT they face. What initially seems like difficult 'multitasking' across different applications at work becomes more the norm as we increase screen real estate and computing power to simultaneously tackle, for example, word processing, web browsing, searching and emailing. What starts off as being 'work life imbalance' becomes 'connectedness', as we check our smartphones more and more frequently, sometimes the last thing before sleep and the first thing in the morning, often from an unknown and difficult to articulate 'fear of missing out' on we know not what. Winston Churchill once said, '*First we shape our buildings and then they shape us*'. In this case, the very features we have designed into ICT, such as reliability, user friendliness and flexibility, are changing the way we use them, and through that, changing our ways of living, working and interacting.

Computers can surely cause us suffering. And yet as history has shown and as is evident from our discussions here, turning back the clock is not an option, neither is endless and perhaps fruitless harping on technologically dystopic futures. It is rather obvious to those that use ICT (which includes most of us) that as new ICT are developed, new effects emerge. What we need to do instead is to remain open to and be courageous about, investigating the unexpected and negative consequences of ICT use in our social, economic and business enterprise, with a view to addressing, mitigating and managing them. It is important and imperative to develop scholarly understanding of such phenomena and understand their organizational effects on efficiency, productivity, innovation and financial performance, and their social effects on relationships, well-being and life habits.

References

Barber, L. K., & Santuzzi, A. M. (2015). Please respond ASAP: Workplace telepressure and employee recovery. *Journal of Occupational Health Psychology, 20*(2), 172–189.

Barley, S. R., Meyerson, D. E., & Grodal, S. (2011). E-mail as a source and symbol of stress. *Organization Science, 22*(4), 887–906.

BBC News. (2012). Volkswagen turns off Blackberry email after work hours. Retrieved March 8, 2012, from http://www.bbc.com/news/technology-16314901.

Beaudry, A., & Pinsonneault, A. (2005). Understanding user responses to information technology: A coping model of user adaptation. *MIS Quarterly, 29*(3), 493–524.

Beaudry, A., & Pinsonneault, A. (2010). The other side of acceptance: studying the direct and indirect effects of emotions on information technology use. *MIS Quarterly*, 689–710.

Beranuy, F. M., Chamarro, L. A., Graner, J. C., & Carbonell, S. X. (2009). Validation of two brief scales for internet addiction and mobile phone problem use. *Psicothema, 21*(3), 480–485.

Bruner, O. (2006). *Playstation nation: Protect your child from video game addiction*. Center Street.

Cooper, C. L., Dewe, P. J., & O'Driscoll, M. P. (2001). *Organizational stress: A review and critique of theory, research, and applications*. Thousand Oaks, CA: SAGE Publications.

Dabbish, L. A., & Kraut, R. E. (2006). Email overload at work: An analysis of factors associated with email strain. In *Proceedings of the 2006 20th anniversary conference on Computer supported cooperative work* (pp. 431–440). Banff, Alberta, Canada: ACM.

Daneback, K., Cooper, A., & Månsson, S.-A. (2005). An internet study of cybersex participants. *Archives of Sexual Behavior, 34*(3), 321–328.

D'Arcy, J., Herath, T., & Shoss, M. K. (2014). Understanding employee responses to stressful information security requirements: A coping perspective. *Journal of Management Information Systems, 31*(2), 285–318.

Davenport, T. H., & Kirby, J. (2016). *Only humans need apply: Winners and losers in the age of smart machines*. Harper Business.

Davies, M. (2015). Checking your emails outside of work really IS bad for your health, 'increasing stress levels and reducing wellbeing'. Retrieved October 8, 2015, from http://www.dailymail.co.uk/health/article-3192467/Checking-emails-outside-work-really-bad-health-increasing-stress-levels-reducing-wellbeing.html.

Day, A., Paquet, S., Scott, N., & Hambley, L. (2012). Perceived information and communication technology (ICT) demands on employee outcomes: The moderating effect of organizational ICT support. *Journal of Occupational Health Psychology, 17*(4), 473–491.

Derks, D., van Duin, D., Tims, M., & Bakker, A. B. (2015). Smartphone use and work–home interference: The moderating role of social norms and employee work engagement. *Journal of Occupational and Organizational Psychology, 88*(1), 155–177.

Eurofound and the International Labour Office. (2017). *Working anytime, anywhere: The effects on the world of work*. Publications Office of the European Union, Luxembourg, and the International Labour Office, Geneva.

Friedman, R. A., & Currall, S. C. (2003). Conflict escalation: Dispute exacerbating elements of e-mail communication. *Human Relations, 56*(11), 1325–1347.

Galluch, P. S., Grover, V., & Thatcher, J. B. (2015). Interrupting the workplace: Examining stressors in an information technology context. *Journal of the Association for Information Systems, 16*(1), 1–47.

Griffiths, M. D., Parke, J., Wood, R., & Rigbye, J. (2010). Online poker gambling in university students: Further findings from an online survey. *International Journal of Mental Health and Addiction, 8*(1), 82–89.

Gupta, A., Sharda, R., & Greve, R. A. (2011). You've got email! Does it really matter to process emails now or later? *Information Systems Frontiers, 13*(5), 637–653.

Huang, H., & Leung, L. (2009). Instant messaging addiction among teenagers in China: Shyness, alienation, and academic performance decrement. *CyberPsychology & Behavior, 12*(6), 675–679.

Hussain, Z., & Griffiths, M. D. (2009). Excessive use of massively multi-player online role-playing games: A pilot study. *International Journal of Mental Health and Addiction, 7*(4), 563.

Iacovelli, A., & Valenti, S. (2009). Internet addiction's effect on likeability and rapport. *Computers in Human Behavior, 25*(2), 439–443.

Jackson, T. W., Burgess, A., & Edwards, J. (2006). A simple approach to improving email communication. *Communications of the ACM, 49*(6), 107–109.

Jackson, T. W., Dawson, R., & Wilson, D. (2001). The cost of email interruption. *Journal of Systems and Information Technology, 5*(1), 81–92.

Jenaro, C., Flores, N., Gómez-Vela, M., González-Gil, F., & Caballo, C. (2007). Problematic internet and cell-phone use: Psychological, behavioral, and health correlates. *Addiction Research & Theory, 15*(3), 309–320.

Kakabadse, N., Porter, G., & Vance, D. (2007). Addicted to technology. *Business Strategy Review, 18*(4), 81–85.

Karr-Wisniewski, P., & Lu, Y. (2010). When more is too much: Operationalizing technology overload and exploring its impact on knowledge worker productivity. *Computers in Human Behavior, 26*(5), 1061–1072.

Kim, E. J., Namkoong, K., Ku, T., & Kim, S. J. (2008). The relationship between online game addiction and aggression, self-control and narcissistic personality traits. *European Psychiatry, 23*(3), 212–218.

Lanaj, K., Hollenbeck, J. R., Ilgen, D. R., Barnes, C. M., & Harmon, S. J. (2013). The double-edged sword of decentralized planning in multiteam systems. *Academy of Management Journal, 56*(3), 735–757.

Maier, C., Laumer, S., Eckhardt, A., & Weitzel, T. (2014). Giving too much social support: Social overload on social networking sites. *European Journal of Information Systems, 24*(5), 1–18.

Maier, C., Laumer, S., Weinert, C., & Weitzel, T. (2015). The effects of technostress and switching stress on discontinued use of social networking services: A study of Facebook use. *Information Systems Journal, 25*(3), 275–308.

Mano, R. S., & Mesch, G. S. (2010). E-mail characteristics, work performance and distress. *Computers in Human Behavior, 26*(1), 61–69.

Mark, G., Voida, S., & Cardello, A. (2012). A pace not dictated by electrons: An empirical study of work without email. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 555–564). ACM.

Matusik, S. F., & Mickel, A. E. (2011). Embracing or embattled by converged mobile devices? Users' experiences with a contemporary connectivity technology. *Human Relations, 64*(8), 1001–1030.

Mazmanian, M., Orlikowski, W. J., & Yates, J. (2005). Crackberries: The social implications of ubiquitous wireless e-mail devices. In C. Sørensen, Y. Yoo, K. Lyytinen & J. DeGross (Eds.), *Designing ubiquitous information environments: Socio-technical issues and challenges* (pp. 337–343). New York: Springer. Retrieved from https://doi.org/10.1007/0-387-28918-6.

McGrath, J. E. (1976). Stress and behavior in organizations. In M. D. Dunnette (Ed.), *Handbook of industrial and organizational psychology*. Chicago: Rand McNally.

Montag, C., Sindermann, C., Becker, B., & Panksepp, J. (2016). An affective neuroscience framework for the molecular study of Internet addiction. *Frontiers in Psychology*, *7*. Retrieved from https://www-ncbi-nlm-nih-gov.bases-doc.univ-lorraine.fr/pmc/articles/PMC5159430/.

Morahan-Martin, J., & Schumacher, P. (2000). Incidence and correlates of pathological Internet use among college students. *Computers in Human Behavior, 16*(1), 13–29.

Niemz, K., Griffiths, M., & Banyard, P. (2005). Prevalence of pathological internet use among university students and correlations with self-esteem, the General Health Questionnaire (GHQ), and disinhibition. *CyberPsychology & Behavior, 8*(6), 562–570.

Ragu-Nathan, T. S., Tarafdar, M., Ragu-Nathan, B. S., & Qiang, Tu. (2008). The consequences of technostress for end users in organizations: Conceptual development and empirical validation. *Information Systems Research, 19*(4), 417–433.

Reinke, K., & Chamorro-Premuzic, T. (2014). When email use gets out of control: Understanding the relationship between personality and email overload and their impact on burnout and work engagement. *Computers in Human Behavior, 36,* 502–509.

Schwartz, M. F., & Southern, S. (2000). Compulsive cybersex: The new tea room. *Sexual Addiction & Compulsivity: The Journal of Treatment and Prevention, 7*(1–2), 127–144.

Soucek, R., & Moser, K. (2010). Coping with information overload in email communication: Evaluation of a training intervention. *Computers in Human Behavior, 26*(6), 1458–1466.

Sprigg, C. A., & Jackson, P. R. (2006). Call centers as lean service environments: Job-related strain and the mediating role of work design. *Journal of Occupational Health Psychology, 11*(2), 197–212.

Srivastava, S. C., Chandra, S., & Shirish, A. (2015). Technostress creators and job outcomes: Theorising the moderating influence of personality traits. *Information Systems Journal, 25*(4), 355–401.

Stich, J.-F., Farley, S., Cooper, C., & Tarafdar, M. (2015). Information and communication technology demands: Outcomes and interventions. *Journal of Organizational Effectiveness: People and Performance, 2*(4), 327–345.

Tarafdar, M., Gupta, A., & Turel, O. (2015a). The dark side of information technology use: Introduction to the special issue and a research framework. *Information Systems Journal*, *25*(3), 161–170.

Tarafdar, M., Pullins, E. B., & Ragu-Nathan, T. S. (2015b). Technostress: Negative effect on performance and possible mitigations. *Information Systems Journal, 25*(2), 103–132.

Tarafdar, M., Qiang, Tu, Ragu-Nathan, B. S., & Ragu-Nathan, T. S. (2007). The impact of technostress on role stress and productivity. *Journal of Management Information Systems, 24*(1), 301–328.

Tarafdar, M., Tu, Q., & Ragu-Nathan, T. S. (2010). Impact of technostress on end-user satisfaction and performance. *Journal of Management Information Systems, 27*(3), 303–334.

Tarafdar, M., Qiang, Tu, Ragu-Nathan, B. S., & Ragu-Nathan, T. S. (2017). https://onlinelibrary.wiley.com/doi/abs/10.1111/isj.12169.

Tracy, R., & Patterson, S. (2014). Fast traders are getting data from SEC seconds early. Retrieved October 29, 2014, from http://www.wsj.com/articles/fast-traders-are-getting-data-from-sec-seconds-early-1414539997.

Turel, O., & Serenko, A. (2010). Is mobile email addiction overlooked? *Communications of the ACM, 53*(5), 41–43.

Turel, O., & Serenko, A. (2012). The benefits and dangers of enjoyment with social networking websites. *European Journal of Information Systems, 21*(5), 512–528.

Turel, O., Serenko, A., & Bontis, N. (2011). Family and work-related consequences of addiction to organizational pervasive technologies. *Information & Management, 48*(2), 88–95.

Vodanovich, S., Sundaram, D., & Myers, M. (2010). Research commentary-digital natives and ubiquitous information systems. *Information Systems Research, 21*(4), 711–723.

Weinstein, A., & Lejoyeux, M. (2010). Internet addiction or excessive internet use. *The American Journal of Drug and Alcohol Abuse, 36*(5), 277–283.

Xu, Z., Turel, O., & Yuan, Y. (2012). Online game addiction among adolescents: Motivation and prevention factors. *European Journal of Information Systems, 21*(3), 321–340.

Young, K. S. (2010). Internet addiction over the decade: A personal look back. *World Psychiatry*, *9*(2), 91–91.

Zuboff, S. (1988). *In the age of the smart machine: The future of work and power*. New York, NY, USA: Basic books.

What Makes a Good Leader: A Tribute to Anup K Sinha

Anamitra Chatterjee

1 Introduction

This chapter is a tribute to Anup K Sinha, his leadership and teaching abilities. I start by providing some perspective to the reader about my interactions with Prof. Sinha in different capacities. This is followed by some reflection on what core human values constitute a good human being and how Prof. Sinha exemplifies them. Literature on leadership, the qualities of a leader and an analysis of Prof. Sinha's attributes follow next. I then move on to discuss the attributes of a good coach and teacher. The chapter concluded with some of my personal reflections.

2 Anup K Sinha and the Law of the Triad

Anup K Sinha had a long and distinguished research and teaching career spanning 10 years at Presidency College, University of Calcutta, and another 25 years at the Indian Institute of Management (IIMC), Calcutta. His previous education in Economics was at Presidency College, and University of Southern California, USA, in the 1970s. As a Professor of Economics, Prof. Sinha is an eminent teacher. Moreover, he is a good human being and above all, a friend. These roles form parts of a differentiated triad.

Prof. Sinha taught me Economics for the first time at Presidency College (1988) and then again for a second time at IIM Calcutta (1994). This was a unique privilege for me, the only such instance in my education life when the teacher, Anup K Sinha, and I, his student, both were brought together by fate, after a gap of 6 years. As the law of 'triad' caught up, after the same gap of 6 years, it was another opportunity

A. Chatterjee (✉)
Ashridge Executive Education Hult, Berkhamsted, UK
e-mail: chatterjee.anamitra@gmail.com

P. Ray et al. (eds.), *Economics, Management and Sustainability*,
https://doi.org/10.1007/978-981-13-1894-8_15

to discuss economics with Anup K Sinha, around my Masters studies and scholarships application at the London School of Economics (LSE, 2000) UK. Three things remained same: The 'teacher, the taught and the City of Joy, Calcutta. A unique triad!

This was a new phase of me knowing him in a different light. I was going back to him to discuss about a 'plunge into unknown' leaving behind my post MBA banking career with Deutsche Bank AG. He was my guide, I would say, a Supervisor, of my scholarship application. He guided me with a lot of care. Prof. Erik De Haan in his book (2012) Supervision in Action refers to 'other qualities and attributes we might hope to find in an 'ideal supervisor' are a passion for learning, flexibility, humour and an ability to see situations from multiple perspectives combined with a sensitivity to the wider organisation issues". He put me at ease, being open and clear about the importance of sharing information with and seeking further guidance from two former teachers, Prof. Mihir K Rakshit & Prof. Amitava Bose.

I also wonder if Anup K Sinha had read Harvard Professor Argyris (1991) 'Teaching Smart People How to Learn', and used it to get his students to learn. I can now look back with admiration how carefully he challenged me to move past my defensive reasoning, and expand my mindset.

As already discussed, I examine a triad of questions in this contribution. I start with examining what values make a person like Anup K Sinha a good human being? This is followed by the question: What makes a good human like Anup K Sinha a good leader? Finally, I examine what makes a good leader like Anup K Sinha a good coach and teacher. Each of the sections that follow addresses a question at a time, followed by some other reflections on the cornerstones for an inquiry into executive coaching practise.

3 Core Human Values

What values make a person a good human being?

The triad of core values that qualify here are credibility, commitment and courage. The concept of credibility of a person's actions and thoughts refers to them offering rational explanations, such that others have faith in the actions of thoughts. Thus, when a person's actions can be trusted, and when those actions are accepted by the community at large, we can call the person credible. Needless to say, the action must be perceived as beneficial to the economic or social interests of the network of individuals that it affects. Commitment refers to the extent to which a person is dedicated to his beliefs and ideas, and what he is willing to do to take what he believes is a good initiative to its logical conclusion. In order to do so, a person must not only have very strong conviction, but also the courage to follow through. At the same time, following through on your conviction results in your credibility. Thus, these core values of credibility, commitment and courage reinforce each other in a virtuous circle.

I have often wondered about Anup K Sinha, and well indeed, he was credible and committed. We now move on to the next question in our triad: how a good person transcends into a good leader.

4 Leadership Triad: A Sum of Three Parts

What makes a good leader? I delved into leadership literature and books. I refer to the definition mentioned below from the book written by an IIMC alumni. I quote from Reflections: Thoughts on future of leadership and business in India, by D. Shivakumar, former Chairman & CEO PepsiCo.: 'Leadership arises, is earned or is bestowed but is sustained through few elements'. It dawned on me that even leadership could be viewed as a sum of three parts: yet another triad!

When we say leadership arises, it is due to the power and position of the role, gained with experience and skills. Thought leadership is the mode through which leadership is earned. Finally, another way to look at leadership being bestowed is through the concept of followership, something that is referred to in executive coaching as honouring leadership. Shivakumar has gone on to talk of the elements required to sustain leadership, but I choose to remain with the triad of power leadership, thought leadership and followership–leadership, as Prof. Sinha is a rare individual who checks most of the boxes right there, and we cannot attribute his leadership to any one of the mechanisms. Before moving on to illustrate what I mean, it is worthwhile to also realise that this definition of leadership holds water regardless of the changing times.

Digital technology is changing information flow, knowledge access, and hierarchies in the organisation. Thought leadership in a democratic and digital information society is both easy and difficult, albeit in separate ways. The digital age enables people to 'follow' others easily through information sharing. The learning and agile mind is necessary for leaders pursuing thought leadership in a digital age. Being a leader in a digital world requires the leader to connect in person, to express ideas succinctly in the digital medium and to build both internal and external followership. Leaders will be expected to be responsive digitally. The triad of power leadership, thought leadership and followership–leadership can flower in a digital world equally, but one can trip easily too, to put it euphemistically!

Examining Shivakumar's definition using Anup K Sinha as the subject, I come to the following conclusion. He did admirably well with the first aspect, that is from leadership that arises as a result of the power and position of the role, gained with experience and skills. We will come to the second aspect later on. With respect to the third aspect, of honouring leadership, Anup K Sinha has had scores of students, colleagues and administrators who unquestionably chose him as their leader on myriad occasions. He is aware of nuanced differences in people's perceptions and is always willing to engage. In the context of 'earning' leadership, through his teaching and engagement with the students around him and other like-minded institutions, he left behind a palpable mark. He was completely immersed in the profession of teaching

economics and building a better working environment for himself and all other faculty colleagues. Perhaps he could have done more in the second aspect, of thought leadership in economics with more quantity of published research and quality books. It is a great loss to people who have not had the opportunity to meet him to not know how his mind worked and through publishing more Prof. Sinha could have extended his circle of influence greatly.

This brings me to the human element and inner world of Anup K Sinha. Did he strive to do more? Was he contented with his own work–life balance and perhaps did not wish to do more, like my favourite teacher and role model, Prof. Dipak Banerjee (DB), Head of Department, Economics, Presidency College, who was also happy working in a 'leisurely' pace.

He ticks two of three boxes in this leadership definition. My contention is based on one human aspect; that Anup K Sinha perhaps preferred to work at his chosen leisurely pace and did not wish to keep up with the twenty-first century rat race of publishing. He was unselfish and helped generations of students. He was not one of those who are leaders in name only (LINO) who often find it difficult to function as leaders, as they are a bit too selfish or too scheming, They continue in leadership roles hurting people, 'love by a thousand cuts' and their institution. Anup K Sinha rose above such LINOs and made a mark in his profession.

All of this assessment is subjective and the discerning reader may feel uncomfortable.

Let me try a more objective approach, using a universally accepted leadership scale. The following are a list of fifteen universal leadership scales, based on the Global Leadership and Organizational Behavior Effectiveness (GLOBE) project, at University of Pennsylvania, USA. A global exercise was conducted to rank the given leadership traits, based on a seven-point scale ranked from the 'most universally desirable' to 'the least universally desirable'. The 'world mean' of select leadership scales (i.e., the average of 61 country means) is provided in the table that follows. Research showed that the top three leadership scales which emerged were Integrity, Visionary and Inspirational.

From our discussion on what makes a good human being it is obvious that Anup K Sinha displayed high levels of integrity, his students across two institutions bear testimony to this fact over the years, and the vast number of people who wanted to contribute to this festschrift volume is evidence of how inspirational he was.

5 What Makes a Good Coach and Teacher?

If we subscribe to the adage that the customer knows best, we can easily conclude that Prof. Sinha was a good teacher. He was feted with the 'best teacher' award multiple times by his students at IIM Calcutta.

Executive coaching helps you visualise your future, not by telling you what to do but by being an excellent listener and bouncing board. An executive coach helps the coachee articulate own goals and then translates them into actionable next steps.

Leadership scale	World mean
Diplomatic	5.49
Inspirational	6.07
Decisive	5.80
Malevolent	1.80
Autonomous	3.85
Visionary	6.02
Autocratic	2.65
Integrity	6.07
Procedural	3.87
Team-integrator	5.88
Face-saver	2.92
Non-participative	2.66
Administratively competent	5.76
Self-centred	2.17
Performance-oriented	6.02

Source Adapted from http://www.inspireimagineinnovate.com/pdf/globesummary-by-michael-h-hoppe.pdf

Anup K Sinha: showed integrity, was a visionary and was inspirational. He ticks all boxes

Would he have also made an Executive Coach? A good Executive Coach can see elements about yourself and the immediate environment you are in that you are blinded to. He is able to ask the right questions just by observing your life, and rather than provide you with a solution he is able to get you to clearly see your options and decisions you can make to see and advance your life. An Executive Coach brings in knowledge and insight, expands your horizons, opens up new possibilities, and empathetically helps build your confidence and skills. A good teacher also has these traits. Many students will vouch that Prof. Sinha has done this and much more for each of them when they have sought his guidance.

6 Reflections of an Executive Coach

I have been working as a practitioner in the field of executive coaching and dealing with various types of leaders and managers. My preference is for 'slow thinking and leisurely pace of working' and its role in executive coaching in a 'fast-paced digital' business world. As an Accredited Executive Coach & Adjunct, Ashridge Executive Education Hult UK, I work extensively with business and academic leaders in India and UK.

"My leadership development and executive coaching practice is informed by own belief in Performance = Potential – (Internal) Interference. I use this belief and the relational coaching approach to support my clients to help them explore and attain

their own goals". I focus on listening, noticing and reflection for my own clarity. I encourage my coachees to introspect, challenge them to identify and reduce their internal interferences. My approach enhances clients' learning agility, the ability to meet oneself and meeting the other, people-led business mind-set & behavioural change".

This brings me to my area of inquiry into my coaching practise. I prefer working at a leisurely pace. I am a slow thinker as explained Daniel Kahneman in his book: Thinking Fast and Slow. Kahneman is the 2002 Nobel Prize winner in Economics. He refers to a 'Systems1' person-thinker, as someone who prefers and does fast, intuitive, emotional thinking and a 'Systems2' person-thinker who prefers and does the opposite, i.e., slow, logical, deliberate thinking. While a person does both, the default is his preference.

I started my coaching practise in Calcutta, India sometime in 2008, when the corporate business environment was not yet digital. We used simple mobile phones to talk and emails to write. No Smartphone, No WhatsApp, No LinkedIn!

Earlier in 2008, I could do my coaching work at my own pace. I am a tad worried now ten years later, in 2018. I wish to inquire into:

1. My preference for 'slow thinking and working at a leisurely pace'
2. Impact of the above on my coaching practise in a fast-paced digital world

Specific issues made me reflect on this as a good topic for own inquiry. I notice that I am slow to make decisions about taking on new work. I also put far more emphasis in terms of what I might be feeling about the business or its people or something.

Kahler (1975) identified five common drivers: be perfect, please others, hurry-up, be strong, try hard that motivate us, and which can also be at the root of some dysfunctional behaviour. These are commonly known as behavioural drivers. Any of these can be rooted in early admonishment from teachers and parents who are seeking to help the child become socially functional, but perhaps do not offer sufficient guidance and praise for adequate behaviour, thus leaving the child overdoing things. In reasonable quantities, these drivers are effective in creating functioning and successful adults. When people do not know when to stop, then dysfunctional behaviour can set in, causing stress and consequent further coping.

I often wonder if this, my low 'hurry-up' driver has the potential to be very unhelpful in a coaching practise if I am spending too much time 'feeling' rather than allowing agility to cope up with the fast pace of few key stakeholders or coaching relationships. I am also interested, from an 'outer arc' perspective to explore what the digital era business people think of an ideal 'work pace' for executive coaches? Is it possible to know? Does the other person have to be 'System1 or Systems2' person-thinker to be able to advise me on that issue and finally, does this matter? If yes, how much is the impact on coaching? What does action inquiry research say about the importance of fast pace for executive coaches and how we can best develop ourselves as coaches? I wish to understand myself better so that this may help me cope with the pace of the digital world.

References

Argyris, C (1991). Teaching smart people how to learn. *Harvard Business Review*, *69*(3), 99–109.
De Haan, E. (2008). Becoming simultaneously thicker and thinner skinned: The inherent conflicts arising in the professional development of coaches. *Personnel Review*, *37*(5), 526–542.
Kahler, T. (1975). Drivers, the key to the process script. *Transactional Analysis Journal, 5,* 3.
Kahneman, D. (2011). *Thinking fast and slow*. Macmillan.
Shivakumar. D (2016*). Reflections: My thoughts on the future of leadership and business in India*. Xavier University Press.