POLITICAL DECENTRALIZATION

Dilip Mookherjee

Department of Economics, Boston University; dilipm@bu.edu

Revised, December 11, 2014

ABSTRACT

We provide a critical overview of the literature on political decentralization. After reviewing first and second generation theories of federalism, we describe recent empirical studies which have focused mainly on determinants of capture and local government accountability emphasized by second generation theories. The essay concludes by describing emerging new issues that deserve more attention in future research: a wider range of political distortions beyond capture and corruption; effects of decentralization on inter-community allocations, learning and experimentation, and on democracy itself; comparing decentralization with other organizational alternatives; issues pertaining to design and implementation raised by first generation federalism theories, such as inter-jurisdictional spillovers, vertical fiscal relations, the range of devolution, and the political economy of implementation.

KEYWORDS: accountability, capture, corruption, democracy, devolution, federalism

1 Department of Economics, 270 Bay State Road, Boston MA 02215; dilipm@bu.edu. This article has been prepared for the Annual Review of Economics. Pranab Bardhan, Jean-Paul Faguet and Stuti Khemani and an Editor provided useful feedback on an earlier draft. I am particularly grateful to Pranab Bardhan for numerous conversations on these topics over many years. This article overlaps partially with Bardhan (2002) and Bardhan and Mookherjee (2013).
# TABLE OF CONTENTS

1. Introduction
2. Conceptual Frameworks and Theoretical Predictions
   2.1 First Generation Federalism
   2.2 Second Generation Federalism
3. Empirical Studies
   3.1 Cross-Country Comparisons
   3.2 Micro Evidence on Intra-Community Capture and Corruption
   3.3 Empirical Studies of Provincial Protectionism in China and Russia
4. Emerging Issues
   4.1 Widening the Conceptual Apparatus
   4.2 Wider Effects of Decentralization
   4.3 Comparing with Other Organizational Alternatives
   4.4 Design Issues Highlighted by First Generation Theories
   4.5 Domain and Extent of Devolution
   4.6 Political Economy and Implementation Issues
1. Introduction

Decentralization of government is a perennially important topic in the discourse on democracy, constitution design and public finance. In the past two decades, there has been renewed interest in decentralization, particularly in the context of developing countries seeking ways to promote accountability of government in the delivery of public services (see, e.g., the 2004 World Development Report). In this article I provide an overview of some of this recent literature, both broader issues generally raised in this area and specific policy issues faced in developing countries. I shall attempt to summarize what we have and have not yet learnt from this literature.

Section 2 starts with a brief overview of the conceptual framework associated with `first generation’ and more recent `second generation’ theories of fiscal federalism, which are distinguished by the focus of the latter on political economy and government accountability. Section 3 summarizes empirical findings from the recent literature on developing countries. The paper concludes with Section 4 which describes a range of questions and policy issues which represent promising directions of emerging and future research.
2. Conceptual Frameworks and Theoretical Predictions

First Generation Federalism

The classic papers in the economics literature on fiscal federalism are Tiebout (1956) and Oates (1972), discussing the benefits and costs of decentralization respectively from the standpoint of efficient resource allocation, while ignoring agency problems within government. Tiebout’s analysis stressed heterogeneity of citizen preferences for local public goods which cause it to be desirable for citizens with disparate preferences to sort into different communities. ‘Voting with one’s feet’ becomes a way of revealing local public good preferences by citizens, which rationalizes decentralization of spending decisions to local governments. The underlying implicit assumption is that central governments cannot efficiently differentiate provision of local public goods across communities. The reasoning underlying this assumption is however not spelt out explicitly. Oates’ analysis describes some of the inefficiencies associated with decentralization: loss of economies of scale, and inter-jurisdictional externalities. Accordingly, the optimal choice trades off the benefits of greater capacity of decentralized governments to cater to heterogeneous tastes, against the loss of scale economies and inter-jurisdictional coordination. The importance of taste heterogeneity relative to the latter costs can vary across contexts and functions, and is ultimately an empirical matter. For instance, countries with greater geographic or ethnic diversity ceteris paribus ought to have more decentralized governments. Central governments
should retain responsibility for nationwide highways or irrigation systems, while local governments can be devolved responsibility for local roads and canals, and so on.

With regard to taxes, it is widely agreed that decentralizing these to local governments would be associated with high inter-jurisdictional externalities --- such as a race for the bottom --- thereby rationalizing centralization of taxes on potentially mobile factors of production. With the exception of real estate taxes which can be imposed by local governments owing to the immobility of property, most other taxes would have to be imposed by upper layers. Consequently the need arises for downward flows of central government revenues to local government. Mechanisms for such vertical fiscal transfers need to be designed, with efficiency and equity objectives in mind, besides incorporating problems of communicating preferences and needs by local to central government. The underlying assumption is that local governments act on behalf of the interests of citizens in their respective jurisdictions, and the latter do not adequately internalize the costs imposed on other jurisdictions by increased resource transfers to their own jurisdiction. This gives rise to classic principal-agent problems in the design of intergovernmental fiscal transfer mechanisms, even in the absence of any political economy or corruption problems within government. In particular, the central government is viewed as a unitary welfare maximizing entity. As a consequence, the role of local governments is limited to reporting local needs to the central government, and administering
grants received in the interests of local residents. In this sense, there is no genuine devolution of authority to local governments in these scenarios.

Second Generation Federalism

Second generation theories of fiscal federalism focus instead on political economy and corruption problems. One strand of this literature focuses on political economy problems within the central government: public good decisions are made by elected legislators, while abstracting from political agency or corruption problems (i.e., by assuming that elected representatives act in the interest of their constituents). Prominent examples are Besley and Coate (2003) and Lockwood (2002), who criticize the assumption of uniformity of centralized public good provision made by the first generation models on conceptual as well as empirical grounds. In their models, centralized provision involve decisions made by minimum winning coalitions among legislators who represent specific districts and do not internalize impacts on other districts. This results in misallocations and uncertainty which undermine the efficiency of centralized allocations, which have to be traded off against inter-jurisdictional spillovers that are not internalized under decentralization. These models generate similar normative conclusions as the first generation literature: the argument for centralization is enhanced when there is less inter-jurisdictional heterogeneity and more spillovers.
The remainder (which forms the bulk) of the second generation literature focuses squarely on problems of agency within government. These problems have become particularly prominent in developing countries in closing decades of the 20th century, as a consequence of centralized development strategies pursued in the Third World during the 1950s and 1960s. Policies were traditionally formulated by planners and politicians in state capitals, and implemented by bureaucrats appointed by and answerable only to them. Very little thought was devoted to ensuring that policies were well-targeted or appropriate to local conditions, or whether the incentives of bureaucrats or politicians were aligned with public interest. Reports of mounting corruption, chronic absenteeism among government service providers, wasteful and poorly designed policies, lobbying and rent-seeking by powerful interest groups provided a growing sense of development efforts going astray, resulting in outcomes that were neither efficient nor equitable. There was a growing sense that these adverse outcomes could be alleviated by devolving many responsibilities to local governments (see, e.g., the 2004 World Development Report).

In the past three decades, mainstream academic literature has witnessed a flowering of formal models of political economy, contracts, corruption and more generally on the role of ‘institutions’ in long-term development. The second generation literature on fiscal federalism rapidly grew as a result of this confluence of policy concerns and availability of analytical tools. Cremer et al (1995) and Seabright (1996) argued that decentralization combined with local democracy could
play a useful disciplinary role in remedying problems of low accountability of
government service providers. Their key assumption is that information concerning
performance of these providers is available only to local citizens in a form that could
not be documented or verified by third parties. This creates a contractual
`incompleteness' between central bureaucrats and their bosses which limits the
monitoring capacity of the latter. A decentralized system wherein services are
provided by locally elected governments allows disgruntled citizens to punish
errant officials by voting them out in subsequent elections. A centralized system
cannot achieve the same discipline even if the politicians are motivated to serve the
public interest, owing to the difficulty of communicating performance information
by local citizens to higher authorities in a form that could be used the latter to
discipline bureaucrats. Moreover, Besley and Case (1995) argued that a comparative
advantage of decentralization is that it permits `yardstick competition', whereby
citizens can use the performance of governments in other jurisdictions as
benchmarks to evaluate their own governments.

In similar vein, Tommasi and Weinschelbaum (1999) used a common agency model
which predicted that the larger the jurisdiction of governments, the higher is the
ratio of principals to officials, resulting in a weakening of connection between pay
and performance. Myerson (2006) stressed the advantages of decentralized
government in providing a screening device for governance talent and a setting for
officials to develop reputations they can exploit subsequently in transiting to offices
in higher levels of government. Case studies comparing performance of irrigation
bureaucracies in India and Korea by Wade (1997) confirmed the superior performance of the decentralized Korean system owing partly to these disciplinary factors. In stark contrast to the recommendations of the first generation literature, these arguments provide a rationale for decentralization on the ground of superior accountability, which apply even in the absence of any significant heterogeneity across jurisdictions.

A related argument in favor of decentralization is made on the basis of competition among different local governments which limits scope for corruption and rent-seeking among government officials. The typical argument of this form due to Brennan and Buchanan (1980) was formalized by Edwards and Keen (1996) and Arikan (2000). The benefits of competition in these models resemble those of Bertrand price competition among rival producers of a homogenous undifferentiated good. It underlies the view expressed by economic historians (Menes (1999, 2003)) that an important contributing factor to declining municipal corruption in the United States during the turn of the 19th century was the expansion of the American frontier and development of railroads, which raised the ‘elasticity’ of the local revenue base to bribe (or tax) rates, thereby reducing the capacity of government officials to extract rents. Qian and Weingast (1997) and Qian and Roland (1998) provide alternative arguments, based on the problem of the soft budget constraint, whereby governments are unable to commit not to bail out financially troubled state-owned enterprises. Competition between local governments to attract private investors raises the opportunity costs of bailouts.
Decentralization can then serve as a useful pre-commitment device which hardens budget constraints for governments. Outcomes of centralization are inferior since central governments are less subject to competition for mobile investors: they compete only vis-à-vis foreign locations.

A number of qualifications and counterarguments to these favorable views of decentralization subsequently began to emerge. Local governments could be more prone to capture by local elites, owing either to relative lack of media attention, or greater capacity of elites to organize at the local level as a result of greater cohesion or homogeneity. These arguments echoed those of designers of the US constitution which provided powers to central governments as a safeguard against the perils of local elite capture (Hamilton, Madison and Jay (1937)). Bardhan and Mookherjee (2000) used a model of political competition with lobbying by special interest groups to argue that it is difficult to assess a priori whether central or local governments would be more vulnerable to elite capture, as there were a number of factors pulling in different directions. In particular elite capture would likely depend on extent of political competition; patterns of political awareness, participation and literacy; local poverty and inequality. These factors are likely to differ across regions, hence the effects of decentralization would be expected to be highly heterogenous, and raise inequality across regions.

In a related critique of decentralization in the specific context of post-transition decentralization in Russia, Blanchard and Shleifer (2000) argued that provincial
governors had been highly susceptible to capture by old industrial firms, resulting in large transfers to these firms and protection from competition from new firms. The central government was less likely to be captured by initial rent holders, being larger in size than local governments, and less directly affected by unemployment implications of closing any particular firm. This view of decentralization causing ‘provincial protectionism’ was subsequently formalized in a theoretical model by Sonin (2003). The model emphasizes the need for economic decentralization to be accompanied by political centralization, i.e., where regional governors cooperate with the federal government, owing to sanctions that the latter could potentially impose on the former. However, Sonin’s model does not address the potential for capture of the federal government itself within a centralized regime.

Treisman (2007) and Cai and Treisman (2004, 2005) provide counterarguments to the view that competition between local governments invariably reduce agency problems. Cai and Treisman (2005) point out that the argument for benign effects of competition depend on the implicit assumption of homogeneity of jurisdictions with regard to the productivity of the factors concerned. With sufficient heterogeneity across jurisdictions the race for mobile factors can be highly uneven, and the worse endowed regions can end up with less business-friendly policies and higher corruption. Arguments based on heightened ‘race-to-the-bottom’ owing to different argument are offered by Cai and Treisman (2004): local governments have incentives to attract investors away from other regions by offering them opportunities to evade central government taxes and regulations, the cost of which
is largely borne by other regions. The broader point is that there are many different ways of attracting investors, some of which are in the public interest (such as reducing rent extraction and providing better infrastructure), and some that are not (such as shielding them from central taxes and regulations). The effects of decentralization will depend on which of these are more effective instruments of attracting private investors. These depend on existing institutions, and can only be settled (if at all) by empirical studies. In this respect, various strands of the theoretical literature on elite capture eventually agree that the effects of decentralization are difficult to predict \textit{a priori}, thereby highlighting the need for empirical analysis.

3. Empirical Studies

3.1 Cross-Country Comparisons

We start by reviewing evidence from cross-country regressions concerning the relation between decentralization and measures of corruption and government performance. Estache and Sinha (1995) reported one of the earliest studies of this nature, based on a panel dataset of 20 countries (comprised equally of developed and developing countries) covering 1970—92, concerning the relation between a measure of expenditure decentralization (the share of sub-national governments in total government spending) and per capita delivery of different forms of infrastructure. They found a significant positive association in general, an effect
which was stronger when there was greater revenue decentralization (measured by
dependence of sub-national governments on self-generated revenues rather than
fiscal transfers). Similar benign (i.e., negative) effects of expenditure
decentralization on measures of corruption (drawn from subjective perceptions of
business-people and investors) were reported by Fisman and Gatti (2002a), using a
dataset covering 59 countries over the period 1980—95. In another paper Fisman
and Gatti (2002b) argue --- analogous to the finding of Estache and Sinha mentioned
above --- that the effects of expenditure decentralization depends on accompanying
revenue decentralization. Using data from 50 states in the US covering the period
1976--87, they show a significant positive association between conviction rates for
abuse of public office and dependence on central transfers, after controlling for state
size, GDP, enforcement expenditures, and civil service wages.

However, Treisman (2002) argues that the Fisman and Gatti (2002a) results are not
robust to choice of measures of corruption and inclusion of additional controls such
as proportion of Protestants in the population (which is positively correlated with
decentralization and negatively with corruption). He shows that expenditure
decentralization tends to be positively correlated with youth illiteracy and access to
paved roads, but negatively with access to sanitation. Appointment decentralization,
measured by the proportion of key officials in sub-national tiers that were directly
elected (or by directly elected local legislatures) rather than appointed by a central
government, exhibited no significant relation with corruption. It had a significant
positive correlation with measures of health services delivered and paved roads, but
also with youth illiteracy rates. The scope for inter-jurisdictional competition --- measured (negatively) by average land area covered by first-tier governments because this serves as a proxy for distance between jurisdictions --- was positively correlated with corruption. Fan et al (2009) find a robust increase in corruption when there are more vertical tiers in government, and when local governments are larger (with size measured by number of employees).

The cross-country studies thus do not provide robust results on effects of decentralization on governance quality. Given the well-known methodological problems with this research methodology --- unobserved cross-sectional heterogeneity, endogeneity bias, data quality and comparability problems --- much of the empirical literature has sought evidence at a more disaggregated level where these problems are less severe.

3.2 Micro Evidence on Intra-Community Capture and Corruption

The past decade has seen a vast outpouring of studies of corruption, elite capture or other measures of governance at the level of individual communities, villages or municipalities. While much of the earlier literature has been summarized in a recent book by Mansuri and Rao (2013), important papers continue to appear more recently. Hence it is worthwhile to update their summary to incorporate recent notable contributions.
The main question continues to be the one posed by Mansuri and Rao at the outset of their book: the importance of ‘capture’ or ‘civil society failure’, defined as the lack of involvement of poor and marginalized groups in decision-making, project outcomes that are less aligned with their needs, capacity of elites to hijack programs to suit their ends, or of officials to pocket rents or divert resources. The key issues are determinants of capture: the studies generally show that capture tends to be higher in communities with greater inequality, those that are “remote from centers of power; have low literacy; are poor; or have significant caste, race or gender disparities” (Mansuri and Rao (2013, p. 5)).

For instance, Galasso and Ravallion (2005) found intra-village targeting of a Food-for-Education program in Bangladesh were less pro-poor in villages that were more remote, and had higher land inequality. Most of the targeting variations were intra-village rather than inter-village in their context, reflecting the effect of actions of local authorities rather than budgetary allocations by higher level governments. Bardhan and Mookherjee (2006a) in contrast found negligible variation in intra-village targeting of services delivered by local governments in the Indian state of West Bengal. On the other hand, they found lower employment generated by local governments out of allocated public work funds, and significantly lower allocation of funds by higher level governments to jurisdictions with higher poverty, land inequality and higher proportion of low caste households within the poor. Using a large household dataset from India, Kochar (2008) found evidence consistent with the hypothesis that landed elites blocked investments in local public education, as
the latter tend to increase wages they paid to hired workers. In Ecuador Araujo et al (2008) found that local communities were significantly less likely to select a pro-poor project in villages with greater inequality. In the Philippines, Labonne and Chase (2009) present evidence of capture by local leaders in highly unequal villages at the project proposal stage; these leaders exercised greater influence over resource allocation at the supra-village level meetings where proposed projects are approved.

In similar vein, Acemoglu, Reed and Robinson (2013) find in Sierra Leone that communities with fewer ruling chiefs (originally recognized by British colonial authorities) have significantly worse development (such as non-agricultural employment, education and child health) outcomes today. They interpret this as a result in reduced political competition among elites who play an important role in allocating land rights within the community. In the Indian state of Maharashtra, Anderson, Francois and Kotwal (2014) show that areas with greater domination of land ownership by the Maratha caste were associated with significantly lower wages, lower take-up of pro-poor government programs, and stronger clientelistic relationships involving insurance and employment between Maratha landlords and the poor. In both Sierra Leone and Maharashtra contexts, greater elite capture was positively correlated with measures of social capital (such as collective action, stated trust in community leaders, and participation in social activities).
Contrasting results concerning the role of elite capture emerge from recent experimental evidence from Indonesia in Alatas et al (2012, 2014). These studies experimentally varied methods for selecting beneficiaries of cash transfers to those below the poverty line, including variants where selections were made only by local elites, or by members of the local community including elites, or were based on data based on community assets. In a low stake experiment, they found no evidence of elite capture (measured by benefits cornered by elites or their relatives, or differences in selection patterns between treatments with varying control by elites). In a higher stake experiment they found some evidence where formal elites holding local government positions were more likely to receive benefits, but the extent of such capture turned out to be quantitatively small.

These studies therefore suggest considerable heterogeneity across and within countries with regard to the extent of local capture. One would then expect the benefits of decentralization to exhibit corresponding heterogenous patterns. Such heterogeneity has been observed in many contexts. For instance, in their study of the distributional effects of decentralization across municipalities on educational quality in Argentine secondary schools, Galiani, Gertler and Schargrodsky (2008) found that schools in poorer municipalities fell further behind, while those in better-off areas improved. In a field experiment (Bjorkman and Svensson (2009)) which randomly varied community involvement in health clinics in Uganda and found higher average levels of service delivery and corresponding improvements in child
health indicators, Bjorkman and Svensson (2010) found these improvements varied significantly with income inequality and ethnic fractionalization.

Given the risk of elite capture, benefits of decentralization are also likely to vary with institutional safeguards against such capture, whether they are accompanied by accountability mechanisms such as contested elections, transparency of budgeting procedures, public provision of information, oversight by citizen councils, judicial authorities, auditors or media. In Uganda, Reinikka and Svensson (2005) found that a newspaper campaign providing school parents with information to monitor local officials’ handling of education grants reduced diversion of these funds substantially, resulting in positive effects on enrollment and student learning. In Brazil, Ferraz and Finan (2008) show that the release of audit reports on local government expenditure of federal transfers had a significant impact on the electoral performance of incumbent mayors. This impact was more pronounced in municipalities where local radio was present to divulge audit information. In Madagascar, Francken et al. (2009) show that the media significantly reduced local capture of public education funds; this effect was stronger in more educated communities which were presumably better able to use the information to monitor local providers. Banerjee et al. (2010) experimented with provision of citizen report cards (evaluating legislator performance and characteristics) to slum residents in Delhi in India and found that the report cards raised the vote shares of the better-performing incumbents. Bjorkman, de Walque and Svensson (2014) describe an experiment with Ugandan health clinics which compared the performance of a
treatment involving community participation combined with information provision (on local staff performance based on household surveys and clinic records), with one consisting of community participation alone. While the former achieved striking improvements in child health outcomes, the latter treatment failed to do so.

An alternative way of repairing civic failures is to mandate the inclusion of disadvantaged groups in the participatory process. A growing body of literature from India, starting with Chattopadhyay and Duflo (2004), has been studying the effects of gender and caste based reservations of mayoral positions in local governments (e.g., Bardhan, Mookherjee and Parra Torrado (2010), Besley et al (2004), Besley, Rao and Pande (2005, 2012)). These papers generally find caste-based reservations resulting in increased targeting of benefits to the corresponding caste groups.

3.3 Empirical Studies of Provincial Protectionism in Russia and China

The studies described in the previous section focus on capture and corruption within specific communities. The theoretical literature has also addressed possible adverse spillover impacts of decentralization on other parts of the country, such as raising provincial protectionism. As explained above, the Blanchard-Shleifer-Sonin viewpoint in this regard contrasts sharply with the Qian-Roland-Weingast market-preserving-federalism hypothesis. What does empirical evidence show?
In the Russian context, evidence in favor of the ‘local-state-corroding-federalism’ hypothesis is provided by Ponomareva and Zhuravskaya (2004) and Slinko et al (2004). Using a sample of Russian firms from the second half of the 1990s, the former show that federal tax arrears accumulated at a faster rate in provinces whose governors had a larger popular base, were in political opposition to the central government, and had a stronger bargaining position. Lambert-Mogiliansky et al (2003) show in a sample of over 8,000 Russian firms that firms located in regions with governors that were politically more powerful, independent from the central government and with more opaque tax collection systems were significantly less likely to be liquidated under the new bankruptcy law of 1998 which enlarged the discretion of regional governors over bankruptcy procedures. Slinko et al (2004) show firms with more preferential treatments grew faster in profits, sales, employment and federal tax arrears after controlling for firm characteristics. Firms without any such treatments invested more and performed better when they were located in regions where political power was less concentrated. Regions with more concentrated power exhibited lower rates of development of small businesses and retail turnover. While overall rates of growth of economic activity in the region were not significantly different, greater concentration of power was associated with lower payment of federal taxes.

China exhibited substantially different results. Jian et al (1999) use data from 1982-92 for 30-odd provinces to show that fiscal decentralization was positively related to growth of per capita regional GDP, of non-agricultural employment and non-state
industrial output, after controlling for provincial tax rates, regional labor force growth, province and time fixed effects. Administrative decentralization had a significant positive correlation with local fixed investment, ratio of local to central government investment, and growth of non-agricultural employment and non-state industrial output. They explain these results as resulting from the effects of altered fiscal contracts between local and central governments, wherein local governments were allowed to retain a higher fraction of tax revenues collected during 1982-92, compared to the previous decade. Despite this, inequality in provincial per capita budgetary revenues and expenditures at the provincial level declined during 1982—92, while inequality in central transfers moved in the opposite direction. They infer that high incentives on the margin co-existed with infra-marginal redistribution in the fiscal contracting mechanism.

A contrasting view of the effects of increased fiscal decentralization in China is presented by Young (2000) and Poncet (2003, 2004), who argue that the increased autonomy and incentives offered to local governments induced them to engage in provincial protectionism, in the form of high barriers to inter-regional trade. Moreover, just as in the Russian case, local governments offered local enterprises exemptions on central taxes, accompanying them with informal levies accruing to their own coffers. They show that the composition of output converged across regions between 1978 and 1997, while prices, labor productivity and labor allocations diverged, and reliance on international trade increased.
Despite the evidence for growing trends of regional protectionism in China, Young (2000, p.1129) concedes that fiscal decentralization probably contributed to economic growth overall, owing to its success in dealing with control and incentive problems. Most writers on China and Russia seem to concur that the overall effects of fiscal decentralization was positive in China, and negative in Russia, and that this owed to a strong central government in China. While this is plausible, there is little hard evidence in favor of this hypothesis. Most of the empirical results for China pertain to data on differences in regional growth rates and their relation to the nature of fiscal contracting: there are no estimates available of the effects of regional protectionism on growth in China as a whole which trade off the superior incentive properties of decentralization against the corroding effects on regional protectionism. The only supporting arguments are provided by Jin et al (1999) who point out that fiscal contracts with the provinces in the two countries were quite different: budgetary constraints for provinces were much `harder' in China. Blanchard and Shleifer (2000) argue that the problem of capture of local governments by local firms was much less severe in China, because China did not start its reforms with the large industrial enterprises that characterized the Russian economy of the early 1990s.

4. Emerging Issues

In this section, I shall describe other important questions that are beginning to receive more attention, and need to be addressed further in future work.
The focus on intra-community allocations and extent to which they reveal mis-targeting, diversion or lack of responsiveness to local need is a natural first step in evaluating the accountability performance of local governments. Much of the existing literature has grappled with the econometric challenge of identifying causal effects of decentralization and of underlying determinants of accountability, using experimental or quasi-experimental methods. A natural concomitant is that effects on a narrow range of outcomes in the short run are evaluated. For instance, patterns of political participation and competition in any given context is taken as given, and viewed as an underlying determinant of the effects of decentralization. The possibility that these may in the long term be affected by decentralization has been largely ignored. Moreover, empirical identification of effects of decentralization policies often reveal very little about the underlying processes at play. Less effort has been devoted to understanding the political economy of local governments in any given context. Consequently when disparate results are found in different contexts, it is difficult to draw any general lessons that could be useful in other areas where decentralization policies are yet to be determined.

These concerns of course apply more to recent empirical literature on policy evaluation more generally: identification of short-term effects on a narrow range of dimensions, concerns about external validity, lack of understanding of underlying processes, and lack of a systemic (e.g., general equilibrium) perspective. In the
specific context of decentralization, the following sets of issues are emerging as important new areas of investigation:

- widening the conceptual apparatus of political economy distortions beyond capture and corruption
- effects on a wider range of relevant dimensions, such as inter-community allocations, and the functioning of local democracy itself (civic participation, political competition, legitimacy, leadership and learning);
- evaluating decentralization vis-à-vis other organizational alternatives
- design issues highlighted by first generation theories of fiscal federalism, such as inter-jurisdictional spillovers, inter-community allocations, and hardness of budget constraints;
- the appropriate domain and extent of devolution; and
- the political economy of implementation.

4.1 Widening the Conceptual Apparatus
While elite capture or corruption are important political distortions associated with local governments, they are not necessarily the only ones. Consider communities where literacy is high, local elections are free and fair by and large, involve high levels of political turnout and awareness across all sections of the population, and are highly contested. Should one then assess the likely welfare benefits of decentralization to be high, on the grounds that such a community is not susceptible to much capture or corruption? Strong political competition could conceivably induce candidates to pander to voters in a variety of ways, by offering short-term benefits (such as loan waivers or low taxes) at the expense of long-term programs.

In developing countries, there is considerable evidence of clientelistic politics or vote-buying, wherein parties offer private transfers or benefits to specific groups of swing-voters conditional on the latter’s voting for them (e.g., Bardhan et al. 2009), Bardhan and Mookherjee (2012), Khemani (2014), Keefer and Khemani (2014), Kitchelt and Wilkinson (2007), Robinson and Verdier (2013), Stokes (2005), Wantchekon (2003)). Such clientelistic relations tend to be more common with poorer voters (as their votes are ‘cheaper’ to purchase). In contexts where clientelism is common, one would therefore observe highly pro-poor targeting. Viewing the outcomes through the singular lens of elite capture, one might infer the outcome to be benign. Yet it could conceal the underlying distortion of clientelism, associated with excess spending on narrowly targeted discretionary private transfers of a recurring nature, at the expense of more broadly targeted programmatic transfer programs or public goods.
In similar vein, mandated reservations of local government positions for disadvantaged minorities defined by ethnicity or tribe is often associated with increased transfers to members of such groups (as much research in the Indian context shows). How such outcomes should be evaluated depend on the relative importance of distortions related to capture and clientelism respectively. Are these increased transfers symptoms of a reduction of elite capture, or heightened clientelism? More attention needs to be devoted to the study of clientelistic distortions, and whether they are heightened by decentralization. Khemani (2013) provides a theoretical model wherein central governments use decentralization as an instrument of dispensing targeting private transfers, at the expense of broad-based public goods. Khemani (2014) subsequently provides evidence from the Philippines of a robust negative correlation between voter reports of having received targeted private transfers, and the delivery of community health services (which in turn was positively correlated with child health outcomes).

In a related vein, politicians and voters are often prone to cash in on short-term benefits at the expense of long term objectives, resulting in problems of credible commitment by politicians to policies that are likely to achieve long-term goals. For instance, there is short-term pressure to avoid levying or collection of taxes and user fees, which undermine the fiscal capacity of governments to make long-term investments in infrastructure. To gain popularity, governments and voters could be prone to spend less on education, health and sanitation the benefits of which are
often diffuse and pushed into the distant future. Is decentralization more or less likely to be vulnerable to such populist distortions? This issue is discussed in a preliminary way in Bardhan and Mookherjee (2013), but it deserves more attention, both theoretically and empirically.

4.2 Wider Effects of Decentralization

Much attention has been devoted to intra-community allocations of government services: how they are allocated between different programs, different categories of beneficiaries, or how efficiently such services have been delivered. In evaluating the relation between underlying determinants of elite capture such as land or socio-economic equality, attention is often focused on a narrow subset of programs administered by local governments (e.g., latrines in Ecuador (Araujo et al (2008)), schools in Argentina (Galiani et al (2008)) or allocation across a few categories of public goods in India (Chattopadhyay and Duflo (2003), Foster and Rosenzweig (2001)). Rarely is a more comprehensive evaluation undertaken which incorporates the larger range of programs. Moreover, effects of decentralization on inter-community allocations are no less important than effects on intra-community allocations. For instance, the 1994 decentralization in Bolivia resulted in a remarkable redistribution of central resources in favor of poorer regions (Faguet (2004)). The Indian state of West Bengal (Bardhan and Mookherjee (2006a)) manifested a bias in the opposite direction, wherein elite capture manifested itself in distorted inter-community rather than intra-community allocations.
The nature of civil society and democracy which has been found to be a key determinant of effects of decentralization in a large number of studies are typically taken as given. Yet, as Faguet (2014) argues in his introduction to a recent symposium on decentralization, there could be important effects on the nature of democracy itself. Citizens may participate more in civic affairs when they perceive more of a stake from doing so. In Indonesia, the 2001 fiscal decentralization was followed by emergence of new leaders that were elected rather than appointed by local elites in about 20% of all villages: these villages manifested superior gains from decentralization (Pal and Roy (2014)). Decentralization could provide room for a new set of political leaders to emerge and prove their mettle, conceivably lowering entry barriers into national politics (Myerson (2014)). Even if resource allocations are unaffected, the fact of wider popular participation in decision-making could generate greater satisfaction among citizens, as Olken (2010) finds in a controlled experiment in Indonesia. This may help enhance the legitimacy of government in the eyes of citizens, and limit threats of political secession. Beaman et al (2009) show that Indian voters’ gender stereotypes regarding political leaders became moderated as they become exposed to local women leaders. Women leaders may accumulate political and administrative experience over time, reducing and even reversing gender gaps in governance performance (Afridi et al (2013)). Decentralization provides the scope for learning from policy experimentation by select jurisdictions, an avenue for spatial diffusion of successful policy innovations,
as exemplified by schooling reforms in Sao Paulo, Brazil (Gemignani and Madeira (2014)).

4.3 Comparisons with Other Organizational Alternatives

Most studies of political decentralization examine performance of local government across different jurisdictions, and/or over time, and relate these to differences in political competition, inequality or social capital. Relatively few studies compare it with centralization, owing to lack of suitable data. Yet that is one of the questions of primary concern. Comparisons with other alternatives frequently arise in policy discussions, such as privatization. For instance, if the provision of an infrastructure, education or health services by a centralized government are considered unsatisfactory, would services improve if responsibility for provision were devolved to local governments? Or would it be better to privatize service delivery, while the government limits its role to financing via budgetary subsidies to private service providers or cash vouchers to households? The trade-offs between decentralization and privatization has not received much attention in the literature, either theoretical or empirical. An exception is Granados and Sanchez (2014), who evaluate the comparative performance of decentralization and privatization in delivery of water services in Colombia in lowering infant mortality rates.
4.4 Design Issues Highlighted by First Generation Theories

Spillovers

Inter-jurisdictional spillovers and loss of scale economies are identified as the principal costs of decentralization in the first generation literature. These have to be traded off against benefits from catering to greater heterogeneity and improved accountability under decentralization. Compared to the latter, there is a paucity of empirical studies that evaluate the former.

Nevertheless, there has been some interesting work on effects of decentralization on environmental quality across political jurisdictions. Sigman (2001) finds evidence of free-riding across countries by comparing water quality in national and international rivers in 49 countries spanning 1979-90. Water quality was significantly poorer in monitoring stations located upstream of international rivers outside the European Union. Using data from 500 river monitoring stations within the United States, Sigman (2005) finds similar evidence of 4% deterioration of water quality in monitoring stations located downstream from states authorized to implement the Clean Water Act. These results obtain after controlling for state, river characteristics and monitoring station fixed effects, with the identification coming from time-varying state authority.
Lipscomb and Mubarak (2007) examine the effects of splitting county boundaries in Brazil using data from 372 monitoring stations located in all eight major river basins. They find a significant increase in pollution close to a river’s downstream exit point out of a newly created county, thereby revealing increased free-riding incentives. Yet, they find no evidence of any effects on overall water quality averaging across all locations. They explain these contrasting results by the fact that splitting counties increased aggregate budgets of county governments, resulting in offsetting increases in spending on sanitation. Burgess et al (2012) provide evidence of increased deforestation in Indonesia resulting from splitting local governments between 2000-08. These results inform the need to correctly evaluate costs of decentralization, besides designing decentralization programs to control trans-boundary spillovers via monitoring and corrective taxes or subsidies.

Vertical Fiscal Relations and Inter-jurisdictional Allocations

Another design issue stressed by the first generation literature are arrangements for fiscal transfers from national to local governments. In many Latin American countries decentralization has been associated with macroeconomic instability associated with inability of central governments to control spending of local government owing to ‘soft budget constraint’ problems (Rodden and Eskeland (2003)). In most other developing countries on the other hand, decentralization has transferred significant spending responsibilities to local governments without
concomitant increases in funds devolved, resulting in complaints of ‘unfunded mandates’.

While the softness of local government budget constraints is partly affected by underlying political economy factors, these could also be affected by the design of vertical fiscal transfer mechanisms. To the extent there is scope for such choices, what are the associated trade-offs? Soft budget constraints can create both problems of controlling fiscal spending and inefficiencies associated with bailouts of improvident localities. On the other hand, hardening budget constraints limits flexibility in inter-jurisdictional allocations in response to changing needs and circumstances, and constrains the ability of local governments to provide essential services.

Fiscal strain can also induce increases in corruption, as shown in an interesting study of Indonesian local governments by Henderson and Kuncoro (2004). They report the results of a survey of 1808 firms in 2001-2002, following a comprehensive nation-wide decentralization. After controlling for various firm and locality specific characteristics, they find that increasing the ratio of property tax revenues retained by local governments to local GDP or central government transfers to local GDP, had a strong negative effect on the number of licenses that local firms had to obtain. One standard deviation increase in either of these variables was associated with 73% and 56% increase in the number of licenses, which in turn was positively correlated with bribes (a doubling of number of
licenses raised the ratio of bribes to firm costs by about .8-.9). They interpret this as the effects of fiscal strain of local governments relative to expenditure responsibilities mandated by the decentralization, which led them to underpay government officials, thereby inducing the latter to rely on bribes to supplement incomes.

A related issue is the extent of discretion retained by upper level governments in reallocation transfers across jurisdictions. The use of formula-based transfers in Bolivia was instrumental in the redistribution of funds in favor of poorer regions observed (Faguet (2004)). In contrast, many Indian states allow upper level governments to retain discretion over these transfers, allowing scope for political manipulation. Gupta and Mukhopadhyay (2014) show that the Rajasthan state government dominated by the Indian National Congress (INC) allocated larger funds for the national employment guarantee program to blocks (a layer of government between the district and village level) where the INC lost the previous election by a narrow margin, and that this helped the INC recover their vote shares. Retaining discretion over inter-jurisdictional transfers enable parties dominating higher level governments to manipulate these to secure political advantages, in a manner rarely visible to citizens. More attention therefore needs to be devoted to the design of interjursidictional transfers and their implications.

4.5 Domain and Extent of Devolution
What responsibilities ought to be devolved to local governments? While authority
over allocations across different local projects and their implementation is devolved
in most decentralization contexts, these can be subject to various mandates and
regulations imposed by upper level governments. Projects may have to be approved
in advance, a time consuming procedure which restricts time available to local
governments to execute projects. There can be restrictions on the use of contractors,
or on the ratio of costs incurred on material procurement. The intent of these
regulations is often to limit scope for local corruption, but in practice can become
instruments of discretionary control exercised by upper level governments to
secure their own political advantage (such as restricting spending by local
governments dominated by rival parties).

Many service areas are often partially decentralized, in the sense that responsibility
is shared between central and local governments. Joanis (2014) provides a
theoretical analysis of the benefits (vertical supply complementarities between
different tiers) and costs (reduced accountability, as voters cannot pinpoint
responsibility for delivery failures) of such arrangements. Despite the pervasiveness
of shared responsibilities, very few empirical studies have attempted to examine
their consequences.

Another important question is the desirability of financial devolution: should local
governments be empowered or authorized to levy taxes on local businesses or
residents? Countries vary considerably with respect to the extent of fiscal autonomy
granted to local governments. As reported in the preceding section, cross-country and country panel studies find that expansionary effects of expenditure decentralization on local services are higher, the greater the extent of revenue-raising devolution that accompanies the decentralization. Bardhan and Mookherjee (2006b) develop a theoretical model which is consistent with this finding, but in which devolution of revenue raising authority could nevertheless reduce welfare. The reason is that with full revenue raising autonomy, local governments subject to substantial elite capture are motivated to impose regressive local taxes where the poor end up paying for services delivered to elites. This problem is avoided by restricting financial devolution. On the other hand, if the local government is not subject to elite capture, financial devolution allows local governments the flexibility to adjust spending in response to variations in local need. Moreover, it is often argued that citizens would be motivated to evade taxes less if they were visibly spent on the jurisdictions in which they reside. Hence there are various pros and cons of altering the extent of financial devolution, which do not seem to have been studied empirically.

The converse question of effects of political devolution in a setting with fiscal decentralization have been studied in the context of China, where a reform requiring some local government leaders to be elected has been introduced in the past two decades. Using a village panel and staggered introduction of local government elections, Martinez-Bravo et al (2014) find that provision of local public goods financed via local taxes doubled, and less land was leased to enterprises resulting in
reduced incomes of prosperous residents and lower inequality. In similar vein, Zhang et al (2004) find that local elections increased share of public investments in local government spending, and a shift of taxes from household to enterprises.

4.6 Political Economy and Implementation Issues

Case studies of recent decentralization experiences of a number of poor and middle income countries show striking differences in design and implementation (Bardhan and Mookherjee (2006c)). In some countries, local governments have independent constitutional authority, while in others they exist at the mercy of upper level governments. Local government leaders can be elected in some countries, and appointed by upper level governments. Even when there is some degree of local democracy, as in China over the past two decades, there can be restrictions imposed on freedom of political parties to form and contest these elections. Most countries, however, impose no such constraints. Elected officials of local governments may be constrained to operate with bureaucrats or service workers appointed by and accountable only to upper level governments in some countries, as in the case of India where most school teachers and health workers cannot be hired or fired by local governments. In others, local governments have far greater control.

Moreover, the extent of decentralization de facto often deviates from what is prescribed de jure. Frequently responsibilities are divided between local and higher level governments, with boundaries that are not clearly demarcated. Officials in local governments often lack requisite administrative skills and training. The extent
of upward and downward accountability mechanisms on local governments can also differ greatly, as can the extent of financial authority or range of functions devolved. As is clear from the evidence described by existing literature, these features have a strong impact on the ultimate outcomes of decentralization.

Are there normative rationales in terms of underlying 'fundamentals' for these cross-country or cross-regional differences, or are they instead the result of the political economy of the process by which decentralization was implemented? This is a question that requires more attention. To do this, one has to ultimately confront the question why central governments willingly devolve power to lower level governments, and the circumstances under which they do so. As country case studies (e.g., in Bardhan and Mookherjee (2006c)) show, historical contexts play a critical role in defining the motive for decentralization, and the shape that it eventually takes. Central authorities devolve power either because they are compelled by threats to their own authority (rising competition from competing parties or regional pressures), or seek to preempt such threats, or owing to crises, external shocks and pressures, or to changes in political ideology, or some combination of these. In some cases (such as Bolivia, Brazil, Indonesia and South Africa) a prior authoritarian political regime was swept away by a new government wedded to an ideology more aligned with decentralization, resulting in a 'big bang' decentralization reform. Such decentralizations were more comprehensive, devolving a larger range of responsibilities to local governments with independent constitutional authority and in a more transparent manner. In countries such as
Uganda or China where strong central authorities faced no challenges, they resorted to administrative rather than political devolution as a way of improving the management of government functions. In India and Pakistan, weak central governments tended to counter rising opposition from regional parties by cultivating local governments. Yet in India, the central government was constitutionally required to seek approval from the regional parties, so was forced to let each state implement the devolution in its own manner. This resulted in enormous heterogeneity in the pattern of devolution across different states in India, in which many state governments implemented central statutes and mandates in ways to hold on to as much power as possible. It is a challenging task to develop a theory of the institutional dynamics of actual decentralizations that can provide a unified account of these disparate experiences.
Literature Cited


Randomized Field Experiment on Community-Based Monitoring in Uganda,”
*Quarterly Journal of Economics,* 735-769.


