Conceptualising Informality: Regulation and Enforcement

Ravi Kanbur*

www.people.cornell.edu/pages/sk145

February, 2009

Contents

1. Introduction
2. Informality in the Literature
3. Towards Coherence: Intervention and Informality
4. The Central Role of Enforcement
5. Conclusion

Abstract

The informality discourse is large, vibrant and expanding fast. But there is a certain conceptual incoherence to the literature. New definitions of informality compete with old definitions leading to a plethora of alternative conceptualisations. While some individual studies may apply a tight definition consistently, the literature as a whole is in a mess. This paper proposes that informality and formality should be seen in direct relation to economic activity in the presence of specified regulation(s). Relative to the regulation(s), four conceptual categories that can help frame the analysis are: (A) regulation applicable and compliant, (B) regulation applicable and non-compliant, (C) regulation non-applicable after adjustment of activity and (D) regulation non-applicable to the activity. Rather than use the generic labels informal/formal, it would be preferable if the analysis focused on these four categories (or even more disaggregated as appropriate). A central determining factor in the impacts of regulation on economic activity across these four categories is the nature and intensity of enforcement. While lack of enforcement is well documented, understanding of its determinants—why and to what extent a government would not enforce a regulation that is has itself passed, and why non-enforcement varies from one context to another, is relatively neglected in the literature. Thus specificity on regulation and on enforcement is the key to achieving conceptual clarity in the analytical literature and in the policy discourse on informality.

* Cornell University: T.H. Lee Professor of World Affairs, International Professor of Applied Economics and Management, and Professor of Economics. This paper is based on my V.V. Giri Memorial Lecture, delivered at Lucknow on 13 December, 2008, to the Golden Jubilee Conference of the Indian Society of Labour Economics. This lecture draws on joint work with Basudeb Guha-Khasnobis and Elinor Ostrom on Informality, and with Arnab Basu and Nancy Chau on Enforcement. I thank them without implicating them in the conclusions I reach in this paper.
1. Introduction

Informality is a term that has the dubious distinction of combining maximum policy importance and political salience with minimal conceptual clarity and coherence in the analytical literature. There is a plethora of definitions, which leads to incoherence in analysis and, at its worst, major policy failures. This paper begins, in Section 2, by establishing this claim. This is followed in Section 3 by an attempt at putting forward a coherent and tightly defined framework for identifying situations that might merit the label “informality.” The proposal is to define informality and formality to relate closely to state intervention and regulation, and indeed to specific interventions. Thus in this approach context is central, so the key is for each study to define carefully which regulation it has in mind, and what it means by informality relative to this regulation. Generic and general definitions, which abound in the literature, are inimical to sound analysis and policy advice. Section 4 takes up a neglected aspect of informality which cannot be ignored once the specific framework proposed here is adopted—the enforcement of regulations. It is argued that that conceptual and empirical analysis of enforcement is an integral part of any analysis of informality, that it is crucial to any policy proposals on intervention and regulation, but that this is a relatively neglected topic in the literature. Section 5 concludes the paper.

2. Informality in the Literature

The informality literature is of course vast. There is a multitude of conceptualisations and definitions.1 Here is how Keith Hart, who is recognised to have introduced the term into the literature more than three decades ago (Hart, 1973), saw informality in a recent review (Hart, 2006, p.25):

“‘The main message of the paper (Hart 1973) was that Accra’s poor were not ‘unemployed’. They worked, often casually, for erratic and generally low returns; but they were definitely working….Following Weber, I argued that the ability to stabilise economic activity within a bureaucratic form made returns more calculable and regular for the workers as well as their bosses. That stability was in turn guaranteed by the state’s laws, which only extended so far into the depths of Ghana’s economy. ‘Formal’ incomes came from regulated economic activities and ‘informal’ incomes, both legal and illegal, lay beyond the scope of regulation. I did not identify the informal economy with a place or a class or even whole persons. Everyone in Accra, but especially the inhabitants of the slum where I lived, tried to combine the two sources of income. Informal opportunities ranged from market gardening and brewing through every kind of trade to gambling, theft and political corruption.”

Thus the multifaceted nature of informality was “present at the creation”, the term attempting to capture a number of features of what Hart observed in Accra. But perhaps the central tenet, also present at the start, is captured in the statement that “‘Formal’ incomes came from regulated economic activities and ‘informal’ incomes, both legal and illegal, lay beyond the scope of regulation.”

---

1 This argument is advanced in Guha-Khasnobis, Kanbur and Ostrom (2006). This section draws on the introduction to that volume.
Hart’s term was taken up rapidly by development studies and by international agencies, by the ILO (1972) in particular, which began to codify the definition of informality, particularly keeping in mind the needs of national statistical authorities in measuring the extent and nature of informality. Two decades on from Hart’s original contribution, ILO (1993, paragraph 5) provided the following definition:

“They [informal enterprises] are private unincorporated enterprises (excluding quasi-corporations), i.e. enterprises owned by individuals or households that are not constituted as separate legal entities independently of their owners, and for which no complete accounts are available that would permit a financial separation of the production activities of the enterprise from the other activities of its owner(s). Private unincorporated enterprises include unincorporated enterprises owned and operated by individual household members or by several members of the same household, as well as unincorporated partnerships and co-operatives formed by members of different households, if they lack complete sets of accounts.”

This enterprise based definition of informality relates to those enterprises that fall outside the purview of laws establishing incorporated enterprises. But they cover a large variety of types of enterprises, and in any event the requirements of incorporation differ across countries which are often based on size of enterprise. Even within a country, statistical systems are not consistent in their definitions. Thus Narayana (2006, p. 93) concludes for India that:

“....in general, differences in concepts and definitions are more than similarities and, hence, the available databases are not comparable for measurement of economic contributions and performance between the formal and informal enterprises.”

The broader literature in the mean time evolved and proliferated definitions of informality. As far back as a quarter of a century ago, barely a decade after Hart (1973), Lipton (1984, pp. 198-201) presented a review of the literature which identified the following characteristics of the “informal sector”:

“.... (1) substantial overlap between providers of capital and providers of labour in each enterprise’; (2) ‘prevalence of perfect, or rather ... near-perfect, competition’; and (3) ‘IS consists largely of “unorganised,” unincorporated enterprises, to which legal restrictions on employment (wage minima, regulations affecting working conditions, etc.) and on acquisitions of non-labour inputs (licences, quotas, etc.) do not apply ....”

In addition to this, many writers evoked smallness of size as a separate feature of informal enterprises.

Further to the above variations, a strand of the literature before and after Hart identifies informality with “instability”, “lack of organisation”, “disorganised.” Elements of this characterisation are present in the literature well before Hart. Geertz’s (1963) contrast between the suq or bazzar on the one hand and “rational enterprise” in the sense of Weber matched similar characterisations even earlier by Lewis (1954) and Boeke (1943). The formal sector is supposed to have, as in Hart’s characterisation above, a degree of rule based stability. It is argued by Guha-Khasnobis, Kanbur and Ostrom (2006), that
this dichotomy has been at the root of major policy failures such as the nationalisation of forests in South Asia, supposedly done to curb deforestation caused by “informal” forestry management. As Ostrom (1990) and Nagendra and Ostrom (2007) have shown, not only was this characterisation wrong, nationalisation led to even greater deforestation through organised corruption. de Soto (2003) for extending the rule of law to the informal sector, although more carefully and prudently made, nevertheless seem to draw from the same strand, or sentiment, in the literature. From a different perspective, Scott (2008), using the metaphor of language, has discussed how “vernaculars” come to be dominated by “universals”, to the detriment of the former.

Most recently, there have been arguments to extend the definition of the informal sector to include not only certain types of enterprises but also certain types of workers as well. To quote Chen (2006, p.76), the object is to:

“...extend the focus to include not only enterprises that are not legally regulated but also employment relationships that are not legally regulated or protected. In brief, the new definition of the ‘informal economy’ focuses on the nature of employment in addition to the characteristics of enterprises.

This new move in defining informality finds its official expression in the recent report of India’s National Commission for Enterprises in the Unorganised Sector (NCFEUS), which uses two definitions, one for the informal sector and one for informal workers (NCFEUS, 2008, p2):

“The informal sector consists of all unincorporated private enterprises owned by individuals or households engaged in the sale and production of goods and services operated on a proprietary or partnership basis and with less than ten total workers....Informal workers consist of those working in the informal sector or households, excluding regular workers with social security benefits by the employers, and the workers in the formal sector without any employment and social security benefits provided by the employers”.

From all of the above, I think it would be fair to say that although individual studies apply tight definitions and consistently, the literature as a whole is a mess of alternative conceptualisations and different measures. Sindzingre (2006, p. 61) summarises a recent review as follows:

““The definitional criteria display logical inconsistencies in terms of hierarchy and exclusiveness. The informal economy is defined via the criterion of a form, that is, a negative form (not being ‘formal’), which coexists, however, with a series of ‘substantive’ criteria that refer to categories and characteristics of firms with variable and non exclusive attributes (e.g., being small firms, urban, unregistered, and so on).”

No wonder, then, that Guha-Khasnobis, Kanbur and Ostrom (2006, pp. 2-3) conclude that:

“Given the prominence of the formal–informal dichotomy in the development discourse, one might expect to see a clear definition of the concepts, consistently applied across the whole range of
theoretical, empirical, and policy analysis. We find no such thing. Instead, it turns out that formal and informal are better thought of as metaphors that conjure up a mental picture of whatever the user has in mind at that time.”

This lack of coherence seems to me to lead to analytical and policy problems. Analytically, information can be used inconsistently across different conceptualisations all of which are linked only by the use of the common term “informality.” Policy analysis can also be affected, since very disparate situations are all given the same label, of “informality”, with a related tendency to apply the same policy instrument to very different situations.

It is perhaps too late to abandon the terms “informal” and “formal”, which have become embedded in the discourse. Rather, I think the answer is for each study to set out precisely what its definition of informality is, and for differences in definitions to be appreciated when analytical or policy results are being derived. The next section develops a framework for such precision.

3. Intervention and Informality

In attempting to put forward a coherent approach to informality, I want to draw on what I think is indeed a common strand in the literature, namely that the distinction between informality and formality is to do with the relationship of economic activity to intervention or regulation by the state. But I want to go further and say that every characterisation of formality and informality needs to specify precisely the regulation concerned. Formality and informality are dichotomies relative only to specific interventions or regulations. This has two immediate implications. First, definitions based on other criteria, such as size, or capital intensity, or degree of organisation, or nature of competition, etc., cannot be used (unless of course the criterion itself appears in the regulation). Second, generic definitions of informality are not be used, or at least are to be treated with careful scrutiny as to their usefulness in analytical and policy discourse.

However, even this tight specification leaves a fair amount of complexity. Imagine a world without intervention and then the introduction of the intervention or regulation. To fix ideas, think of this as being a minimum wage regulation for enterprises above a certain size in a particular sector. Economic agents have to decide simultaneously whether to comply and how to adjust to this intervention. This creates several possibilities for agents and their activities post-intervention, relative to the pre-intervention situation:

A. Stay within the ambit of the regulation and comply.

B. Stay within the ambit of the regulation but not comply.

C. Adjust activity to move out of the ambit of the regulation.

D. Outside the ambit of the regulation in the first place, so no need to adjust.
If we had to use the labels “formal” and “informal”, category A would perhaps merit the label “formal”. If the rest of the categories are “informal”, then B is clearly “informal-illegal.” What about categories C and D? For both of these categories, the regulation does not apply—for example, because the regulation stipulates minimum enterprise size, and these enterprises are below that minimum size. But there is clearly a difference between them—category C has adjusted size to come below the minimum size, while category D was below the minimum size in any case so the regulation has not affected it at all. Should either, neither, or both, of these be labeled “informal”? To pose the question is to reveal the problem that a single label in this case will obscure more than it reveals. A, B, C and D are distinct categories with specific economic features in relation to the regulation under consideration. It is best to keep them separate unless a compelling case can be made that aggregation is analytically revealing and provides for better policy analysis.

Of course the above exercise can be done for a set of regulations—for example for minimum wage laws, social security provision, health and safety regulations, etc., across a broad range of sectors. As the set of regulations expands, then perhaps the size of category D shrinks as fewer and fewer activities stay outside the intended ambit of the regulations as a group. Indeed, the size of category D is sometimes taken as a measure of the extent of “the problem” if the regulation is taken to be “good”, since this is the group of agents who are not covered by regulation. But again, generality and genericness is problematic. I would argue that it is better to consider the issue regulation by regulation, and conduct a detailed analysis of the impacts of each, since these may be very different within the categories A, B, C and D, and even more so in relation to movements of activity between these categories.

None of the above is to gainsay the argument that it is useful to have broad numbers on the general characteristics of an economy. For this, generating numbers of workers who are or are not covered by a broad set of regulations might be useful. But then one has to be careful that these numbers are comparable across countries and over time, which means specifying the regulations in the set, and further clarifying whether two regulations with the same label, in two different countries or at two points in time, are in fact the same. Or, indeed, even if the labels do mean the same thing, the extent to which even the same formal regulation is enforced differently in different situations. This leads me to a discussion of what I consider to be a relatively neglected topic in the conceptualisation of informality.

4. The Central Role of Enforcement

The economic agent faces a decision which takes into account the costs and benefits of (i) staying or not staying within the ambit of the regulation and (ii) if staying, then complying or not complying with the regulation. But it should be noted that the sequence runs from (ii) to (i). The way to think about this is that conditional on staying the agent decides whether it is better or not to comply, and then, given this conditional decision, whether it is better to stay or leave the ambit of the regulation. Thus, for example, if enforcement were perfect then the category B would be empty and the only choice would be between category A and C. But with less than perfect enforcement and compliance
all three categories A, B and C are affected. The intensity of enforcement affects decisions which permeate through the different categories—in particular, how many and which sorts of activities will be found in each category. Thus understanding enforcement is central to understanding the nature and character of formality and informality.

That the enforcement of regulations in developing countries is less than perfect is well established. Taking the case of labor regulations, for example, there is a large literature that shows significant violations the world over. For minimum wage legislation, for example, non-compliance is found in Brazil (Lemos, 2004), Costa Rica (Gindling and Terrell, 1995), Honduras (Gindling and Terrell, 2006) and Indonesia (Harrison and Scorse, 2004). For Mexico Levy (2007, p 88) argues that:

“….illegal salaried labor is substantial: 36.5 percent of total nonpublic salaried labor. In other words, in the absence of evasion, [social security] coverage would be at least 58 percent higher than it is today!”

For India, non-enforcement of labor regulations is well known. Here is how Papola, Mehta and Abraham (2008, p 12) summarise two studies for India:

“...in Gujarat, Unni(2000) found that out of the 53 branches of employment listed for fixation of minimum wages by the Gujarat Government in the informal sector….a majority of workers, especially women, earned less than they were entitled to. Patel (1990) in a study of sacked mill workers who have found employment mostly in the informal sector earned what was below the level of the statutory minimum wage.”

There are many reports of this type for India. For example, Madheswaran, Rajasekhar and Gayathri Devi (2005) document non-enforcement for beedi workers in Karnataka. For Maharashtra, Sundar (2008) also documents non-enforcement of a raft of labour regulations. For India as a whole, Papola, Pais and Sahu (2008) show widespread prevalence of wages below official minimum wages.

The nature of labour inspection is described by Papola, Mehta and Abraham (2008, p. 13) as follows:

“The inspectors are reported to meet their quota of inspection by inspecting along easily accessible locations. Quite often they fill the forms themselves sitting in the office. Employers on such cases have a field day in infringing the law by paying less than the minimum wages fixed for their workers. This is evident from the statistics on actual wages vis-a-vis the statutorily fixed minimum wages, as reported in several studies.”

Thus there is widespread lack of enforcement of regulations. But this raises three important questions. First, this lack of enforcement is not uniform over time and across space—what accounts for the variations? Second, why should a government pass a regulation which it then does not enforce? Third, how exactly should we take into account non-enforcement in discussing the pros and cons of new (or even old) regulations, since non-enforcement will create illegality and therefore informality? These, I would argue, are still open questions in the literature, after more than three decades of debate and discourse on informality.
These three questions all point to one thing—we need a conceptualisation of enforcement and the enforcement process. We need a theory of why governments enforce to differing extents. We then need to embed this theory into the evaluation of regulations. Only then will we have a complete picture of informality and its determinants. Such theory, such conceptualisation, is relatively speaking neglected in labor economics. It certainly seems to be neglected in Indian labour economics literature, where the focus has been much more on documenting non-enforcement rather than following through on its implications. There is a theoretical literature that can help us get started. For example, the early work of Ashenfelter and Smith (1979) and later work of Yaniv (2001) both address the issue of the government’s enforcement decision when enforcement is costly. Basu, Chau and Kanbur (forthcoming) consider optimal government choice of the minimum age and enforcement intensity as a joint decision in the two policy instruments, and furthermore pay special attention to the issue of the credibility of enforcement. It is all very well for the government to claim it is going to enforce, and even have sufficient inspectors to enforce, but if the inspectors “turn a blind eye” and are known to turn a blind eye, the government’s claim will ring hollow (especially if the inspectors are following the government’s own implicit rather than explicit, lead. The theoretical concerns modeled in Basu, Chau and Kanbur (forthcoming) are present in the policy context in India in the report of the Second National Labour Commission (Government of India, 2002, paragraph 12.250):

“In West Bengal, when we enquired why the minimum wage law was not being enforced, we were told that both the Trade Unions and the Government Department had agreed to the below-minimum wage payments as both were agreed on protecting the jobs of beedi workers. We have enough reasons to believe that similar arrangements are entered into elsewhere too by the enforcing authorities and the representatives of workers. We believe that any law that creates such a situation becomes a mockery, if not a self-inflicted fraud. We, therefore, feel that we should legislate only what is capable of being put into practice at the ground level. Anything higher that is desirable will have to remain an aspiration or an eventual goal, not a clause in the law. Any other course will breed disrespect, unconcern and contempt for the law and law enforcing authorities.”

So how can the government convince the market that it is serious about enforcement, if it wants to do so? The theoretical results in Basu, Chau and Kanbur (forthcoming) argue that one way of doing this is to have a high official minimum wage. If the government’s “loss function” depends on the gap between the official wage and the actual wages ruling in the market, by legislating a higher minimum wage it can signal credibly its intention to enforce more and thus narrow this gap. This is in fact an argument put forward by ground level activists like the Self Employed Women’s Association (SEWA) for why they lobby for higher official wages even in the face of non-enforcement—because with a bigger gap between official and actual wages, they can then lobby for more enforcement effort, with a resulting higher actual wage. The theory of Basu, Chau and Kanbur (forthcoming) captures this ground level strategy of activists working for the poor.

The key conceptual point is that the optimal level of the minimum wage cannot be seen in isolation from its enforcement; the two are jointly determined and in turn determine who complies, who does not, and who moves out of the ambit of the regulation—in other words, categories A, B and C. Further, attempts to extend the regulation to category D will create new members of categories A, B and C—the precise nature of this allocation will depend on the precise nature of the enforcement problem. This applies of course to any regulation and the allocations it creates across these categories. The conceptualising of informality is thus intimately connected to the conceptualising of enforcement.
5. Conclusion

The informality discourse is large, vibrant and expanding fast. But there is a certain conceptual incoherence to the literature. New definitions of informality compete with old definitions leading to a plethora of alternative conceptualisations. While some individual studies may apply a tight definition consistently, the literature as a whole is in a mess. This paper proposes that informality and formality should be seen in direct relation to economic activity in the presence of specified regulation(s). Relative to the regulation(s), four conceptual categories that can help frame the analysis are: (A) regulation applicable and compliant, (B) regulation applicable and non-compliant, (C) regulation non-applicable after adjustment of activity and (D) regulation non-applicable to the activity. Rather than use the generic labels informal/formal, it would be preferable if the analysis focused on these four categories (or even more disaggregated as appropriate). A central determining factor in the impacts of regulation on economic activity across these four categories is the nature and intensity of enforcement. While lack of enforcement is well documented, understanding of its determinants—why and to what extent a government would not enforce a regulation that is has itself passed, and why non-enforcement varies from one context to another, is relatively neglected in the literature. Thus specificity on regulation and on enforcement is the key to achieving conceptual clarity in the analytical literature and in the policy discourse on informality.
References


