5

The Relevance of the Concepts of Formality and Informality: A Theoretical Appraisal

Alice Sindzingre

5.1 INTRODUCTION

Since its first appearance in the 1970s, the concept of informality has referred to highly heterogeneous phenomena and measurement methods, and has been explored by development microeconomics and institutional economics via the notions of formal and informal institutions and contracts (Hart 1973). This plurality of meanings and instruments calls into question the concept’s validity. The paper synthesizes the critical issues within both a development economics and an institutional economics perspective. It highlights the continuity of ‘formal’ and ‘informal’ phenomena as well as the weaknesses of most criteria that discriminate between the two categories. It proposes a theoretical approach toward institutions according to which all institutions, be they formal or informal, include forms and contents. The latter’s dynamics are explained in terms of core attributes of institutions and rules: relevance, credibility and
enforceability. This approach enables one to understand the impact, resilience and transformation of institutions and the links between them.

The following section examines the conceptual problems inherent in the concept of informality, especially the plurality and heterogeneity of the phenomena and of measurement. It highlights definitional weaknesses, the logical inconsistencies of the formal-informal dualism, and the continuity between the two categories. From an institutional perspective, the third section shows the difficulty in establishing criteria of formal and informal institutions and contracts. Criteria such as the credibility of rules and enforcement capacity have greater explanatory power. Together with the distinction between the forms and contents of institutions, these features more accurately explain the structure and transformation of activities and institutions, which are described through the formal-informal duality.

5.2 THE CONCEPT OF INFORMALITY: HETEROGENEITY AND INCONSISTENCIES

The term ‘informal’ was coined by Keith Hart in his article on informal income opportunities in Ghana, while the 1972 ILO report on employment and poverty in Kenya was the starting point of the subsequent notoriety of the ‘informal sector’.²

5.2.1 Heterogeneous Phenomena and Dualisms

The literature highlights the vagueness and plurality of definitions of the informal economy³ as well as the difficulty in measuring phenomena that are not well defined, as
shown by the wide range of terms: non-observed, irregular, unofficial, second, hidden, shadow, parallel, subterranean, informal, cash economy, black market, unmeasured, unrecorded, untaxed, non-structured, petty production, and unorganized.

Over time the informal sector has come to refer to increasingly heterogeneous phenomena. The informal sector may sometimes be distinguished from the informal economy. It refers to unregulated labour-intensive activities, self-employed entrepreneurs, micro and small enterprises, activities that take place outside state regulations or formal firms, unregistered activities, or which include various degrees of illegality, such as escaping taxation, non-compliance with labour regulations, financial transactions outside the monitoring of the state or banks (from capital flight to petty village moneylenders), and smuggling. Casual work and domestic labour may or may not also be considered as belonging to the informal sector. The multiplicity of definitions and lack of clear conceptual boundaries make it difficult to view these activities as forming a ‘sector’.

The formal-informal dualism is sometimes confused with other dualisms, for example, state-non state, public-private, large-small firms, rigid-flexible norms, market-non market institutions, Western origin-‘traditional’, written-unwritten rules, impersonal-personal and efficient-inefficient enforcement, the latter relying on personal exchanges and mechanisms of trust and reputation. The informal-formal dualism is also associated with the core model of economic dualism, that is, the dualistic nature of labour markets in developing countries. The informal sector is a residual of distortions pushing wages above equilibrium level and hence workers into the unprotected informal sector. This two-sector model (industrial vs. agricultural, capitalists vs. farmers) was elaborated by Arthur Lewis in his analysis of the labour surplus moving from
agriculture to industry and urban jobs (the unlimited supply of rural labour as a condition of industrialization). The informal sector refers to the non-agricultural and low-productivity sector where the unskilled are in excess supply vis-à-vis demand.

5.2.2 Plurality and Heterogeneity of Measurement Methods

Definitional problems also stem from the plurality of methodologies aimed at quantifying informal activities. There is no unique statistical aggregate that corresponds to the concept of informal economy. Statistical certainty is limited to the sub-sectors, such as the types of enterprises or employment that fulfil certain criteria of size, organization, payment of particular taxes, and so on.

The quantification of the phenomena assimilated into the concept of unobserved economy is deduced through various tools that produce large approximations: national accounts, macro-models—based on monetary flows or stocks, indicators, discrepancies, or sets of explanatory variables (e.g., tax, regulation)—and household and enterprise surveys. These quantification exercises rely on several definitions (e.g., underground, informal) and methods that cannot be added up so as to provide a rigorous idea of the economic extent of the ‘informal sector’. The plurality of measurement methods leads to large variations in estimates as well as figures that are both high and highly aggregated. For example, the informal economy (‘unreported income from the production of legal goods and services’) in 2000 would have thus represented 41 per cent of the GNI in developing countries (Schneider 2002; Schneider and Enste 2000). Likewise, informal employment would have represented between one-half to three-
quarters of non-agricultural employment in developing countries, but the figures are higher if informal employment in agriculture is included.\textsuperscript{10}

\textbf{5.2.3 The Logical Inconsistencies of the Definitional Criteria}

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). The 15th ICLS\textsuperscript{11} Resolution (1993), which is the source of the system of national accounts (SNA93) definition, defines the informal sector as ‘units engaged in production’, ‘generating employment and incomes’, displaying a ‘low level of organisation’, and relying on ‘casual employment’, ‘personal and social relations’, ‘rather than contractual arrangements with formal guarantees’.\textsuperscript{12} Here a (‘negative’) form is at the same time defined as a set of features.

Another logical problem is that informal activities may be defined as a sector (manufacturing or services). They also refer to firm features (small-scale, family or self-employment-based, low productivity, low barriers to entry, and low skills) as well as to employment. The ILO recognizes that the enterprise-based definition of the 15th ICLS cannot apprehend informal employment (Hussmanns 2004a). The informal economy refers to a type of activity for the ILO, that is, employment in the informal sector plus informal employment outside of the informal sector.\textsuperscript{13} The levels of the sector and the firm differ, as firms may be heterogeneous (both formal and informal).
The formal-informal dualism may lead to trivial observations, as it represents a partition of the world in two categories. Both categories are defined by default and are residuals of each other. These logical problems are a dimension of measurement problems. For example, one of the large surveys of the informal sector in Niger (1987-88) found that 99 per cent of the enterprises were informal, as the informal sector was defined by default—all enterprises that did not pay taxes on profits (Augeraud 1991).

5.2.4 Concepts that stem from Measurement Objectives

Concepts logically pre-exist vis-à-vis the methods that are devised for testing them. In the case of the informal sector, methods often construct the concept, which has contributed to the plurality of its meanings. There is a discrepancy between the informal sector as an artefact of national accounts with specific criteria (e.g., the ‘non-observed’ economy), which is coherent with other macroeconomic aggregates and allows for international comparisons, and the plurality of its meanings at the microeconomic level. At the microeconomic level, the targets of household, employment and small enterprises surveys are not necessarily homogeneous with those of national accounts and macro-models (e.g., understanding the dynamics of micro-enterprises).

Consensual definitions have stemmed from national accountants’ objectives of measurement, such as achieving exhaustive and comparable estimates of GDP and defining economic production according to the SNA93. Informal sector production is defined as the ‘productive activities conducted by unincorporated enterprises in the household sector that are unregistered’. The 1993 ICLS on which the SNA93 is based has recognized the difficulties of merging a political and analytical concept with a
statistical concept. It highlights the two possible definitions of the informal sector: through the relationship with the legal and administrative system, where informality is equated with non-registration, or as a specific form of production, which contradicts the first definition, as this form of production may include registered firms.

5.2.5 Concepts Determined by Policy Objectives

The concept of ‘informal’ has been significantly influenced by policy and operationalization objectives. National accounts are a matter of national interest and are politically sensitive. Due to their importance in many developing countries—particularly in low-income and aid-dependent countries—, international organizations have had a crucial role in the expansion and shaping of the concept. They have given rise to studies of the informal sector that were determined by political and policy goals. The meanings of the ‘informal sector’ have been influenced by thinking as to how it reacts to economic reform and its possible roles in policymaking—for example, the training of small-scale enterprises or the impact of government regulation and ‘business climate’.

In Keith Hart’s study, informal activities were analysed as multiple opportunities for increasing the income of the poor. The focus on dualism was developed later, for example with the ILO. The Bretton Woods institutions had an interest in the informal sector because the accuracy of the calculations of GDPs and growth, and the strengthening of the capacities of the statistical services. Taxation has also been a key issue for the IMF in the context of stabilization programmes. The informal sector was viewed as a factor in the low level of revenue and has been the subject of various reform
proposals—for example, presumptive taxation or taxation of production factors (Taube and Tadesse 1996).

5.2.6 Non Discriminating Characteristics: Registration, Barriers to Entry, Income

The features of informality do not constitute necessary or sufficient characteristics that would unambiguously assign phenomena in either the formal or informal category. Even the labour statisticians of the 15th ICLS have recommended avoiding a dualist segmentation of the economy and employment: certain activities that are excluded from the informal sector are not necessarily formal, such as domestic unpaid services or small-scale agriculture. For the ILO, the informal economy contrasts with the formal economy, but also with the criminal economy and the ‘reproductive or care’ economy (ILO 2002b).

The continuity between economic activities may be demonstrated in opposition to the hypothesis of segmentation. The apparent discontinuity may be an endogenous response to heavy state regulation and should disappear if the latter were to be relaxed (De Soto 1989; Djankov et al. 2003). There are numerous linkages between formal (public and private) and informal markets: for example, formal firms hiring workers via informal sub-contracting, informal firms using inputs produced in the formal economy and informal lenders applying for bank credit (Aryeetey et al. 1996). Furthermore, individual multiactivity is a dimension of poverty and uncertain environments. In Russia, labour markets display the three statuses of formality, informality, and multiactivity, the most precarious situation being to possess only one formal job (Kalugina and Najman 2002). Individuals may implement rational risk-mitigating
strategies and shift from the formal to the informal and vice-versa or manage parallel 
activities.\textsuperscript{18} Households are heterogeneous and may include members who work in both the formal and the informal sector.

A supposed characteristic of the informal sector is the unrecorded character of its activities. The latter is, however, more a matter of a continuum than a binary dichotomy. Informal firms are often registered by state services (at least a segment of their activity) and pay certain taxes (licenses or taxes on equipment).\textsuperscript{19} Low barriers to entry are supposed to characterize the informal sector, as well as competitiveness and being unregulated. Barriers to entry may indeed be features of the informal sector, created, for example, by a lack of skills, of credit\textsuperscript{20} or of ‘informal’ networks and similar trust-building devices, which may be highly structured in accordance with complex social rules.\textsuperscript{21} The informal sector may therefore be a fragmented market and split between insiders and outsiders (Bardhan and Udry 1999). Differences are more complex than a simple dichotomy. Informal firms may be less capital-intensive, have little access to credit, with heads being less educated but they may be efficient.\textsuperscript{22}

Finally, the informal sector is often characterized by lower levels of income relative to the formal sector—informality being a ‘survival mechanism’. The informal sector, however, displays large variations in earnings. Informal enterprise heads may have higher earnings than the average wage in the formal sector, but employees may earn less than the official minimum wage. In terms of wages and working conditions, workers in the informal sector may be in a worse situation that informal entrepreneurs and workers in the formal sector.\textsuperscript{23} The return on skills is more of a determining factor in explaining
wages differentials and unemployment, and the dichotomies skilled-unskilled and employed-unemployed may be more pertinent than formal-informal (Teal 2000).

5.3 DEEPENING THE INSTITUTIONAL PERSPECTIVE

The continuity between ‘formal’ and ‘informal’ phenomena also operates at the level of institutions. The analysis of institutions suggests a more accurate dichotomy that distinguishes the forms and contents of institutions and their respective dynamics. The focus on institutions shows that instead of the formal vs. informal dichotomy, the phenomena gathered under the concept of informality may be explained by other characteristics, in particular, the credibility and enforceability of institutions and contracts, be they formal or informal.

5.3.1 Informal Institutions, Contracts, Norms: Confusing Characteristics

The concept of informality has been explored by Douglass North in his analyses of institutional change (North 1990; 1991). As is well-known, North defines institutions as constraints that structure political, economic and social interactions and consist of informal—that is, self-enforcing—constraints (sanctions, taboos, customs, traditions, codes of conduct, conventions, norms of behaviour) and formal regulations (constitutions, laws, property rights). Formal constraints are ‘created’, written and intentional, whereas informal constraints evolve over time and are unwritten. The role of institutions is to reduce uncertainty, introduce regularity and stability by establishing a stable (but not necessarily efficient) structure to human interaction.
The concept of informality has also been used in development microeconomics with notions such as informal exchanges, transactions, norms, contracts, interlinked arrangements, determined by informational constraints, and operating in fragmented markets (credit, insurance, labour, land). They underlie, for example, tied labour, sharecropping, insurance, and risk-pooling mechanisms. Informal exchanges and contracts are contrasted with formal ones, that is, written and guaranteed by a state, and are therefore implicitly equated with unwritten and ‘traditional’ ones. Informal arrangements, transactions, and contracts are supported by mechanisms such as personal and repeated exchanges, trust, and reputation, in environments that are characterized by uncertainty, problems of collective action and free-riding.

‘Formal’ contracts are therefore equated with ‘written’ ones and contrasted with ‘informal’ contracts, ‘unwritten’ and ‘outside the state’s sphere’, though these equivalences are rarely theoretically justified. Whether formal or informal, features such as asymmetric information or problems of incentives are more pertinent in explaining the functioning of institutions and contracts. To argue that because a rule is written its nature changes accordingly requires a theory of the cognitive effects of writing on human thought and behaviour. Trust may also be viewed as an indispensable element in the efficiency of formal (written) contracts, as it reduces transaction costs and improves credibility. Moreover, informal transactions may stem from contracts but also from social norms that are by definition not voluntary, for example redistribution within social networks: these transactions constitute ‘informal insurance’, but this is only one of their functions—and a function is not the cause of a norm. These norms may also entail no reciprocity. Economic networks may depend on ‘fundamental social relationships’ between individuals (Udry and Conley 2004). Pre-existing social statuses
(through kinship, ethnicity) may ease trust rather than trust easing informal transactions. Informal transactions may be the actualization of pre-existing statuses, which ex ante include exchange obligations (in labour, goods and individuals) (Platteau 1994). Likewise, informal contracts may efficiently organize market transactions as well as non-market ones.

Finally, the equation of formal institutions with written, intentional creations often results in equating them with complex institutions. This property of complexity is not discriminating and informal institutions also include highly complex rules. These problems of conceptual boundaries also explain the frequently inconclusive character of models that include informal phenomena as one of their variables (often based on very simple indicators) (Sindzingre 2004).

5.3.2 Cognitive Foundations of Institutions: The Issue of Credibility

Douglass North has recognized that informal rules were ‘extensions, elaborations and modifications’ of formal rules and formed a continuum. Individuals make permanent cognitive tradeoffs between multiple norms from various origins (state, kinship, and so on) and assign relevance and credibility to given institutions according to their assessment of situations. Individuals are rational in the sense that they minimize the cognitive costs of mental processes given their limited information (North 1990: 40). For individuals, the fact that rules are ‘formal’, written or coming from the state (or do not) does not determine their credibility in a particular situation and is no more relevant than the many other attributes that orient individuals’ tradeoffs regarding compliance with it: especially rule’s credibility and the benefits or costs of complying with it. State
rules and institutions may be less credible in certain developing countries than local institutions (e.g., social insurance networks). When individuals are exposed to competing norms (‘modern’ and ‘traditional’), tradeoffs are determined more by the relevance and credibility of the norms, and hence by historical and cognitive path dependence. Credibility is also a key property of contracts, as the notion of contract entails a capacity to commit, whether the contract is formal or informal.

Douglass North rightly insisted on the cognitive dimension of institutions (‘cognitive institutionalism’) and on mental models in the shaping of institutions. Learning processes, however, are more determining for North than relevance and credibility. He maintains the distinction between formal (result of human design) and informal institutions (result of ‘spontaneous interaction’) (Mantzavinos et al. 2003: 7). These formal-informal and cognitive distinctions cannot be assimilated, however. Informal institutions, even in North’s sense of traditions, also result from human design. Furthermore, these distinctions do not explain the core question of institutional transformation, that is, why some institutions and rules are more resilient, credible, and complied with than others.

The state is supposed to constitute a discriminating element. Norms and contracts are said to be formal not only because they are written, but also because they are guaranteed by a governmental legal system, in contrast with norms that are maintained ‘privately’ by social groups (e.g., traditions) through personal transactions and reputational mechanisms. This distinction, however, is not convincing. In developed countries, contracts are also supported by trust and cooperative and reputational self-reinforcing mechanisms in parallel with government-enforced laws (Seabright 2004). ‘Informal’ cannot be equated with ‘trust’ and ‘formal’ with ‘written and guaranteed by the state’.

5-13
In many developing countries states are undermined by problems of political legitimacy and institutional credibility.\(^{28}\) Formal contracts may be guaranteed by a state legal system but may be less credible than ‘informal’ contracts guaranteed by other sources of authority built over time.\(^{29}\)

In the perspective of methodological individualism—that is, if the acceptance of rules is considered to be primarily a psychological phenomenon—there is a cognitive continuity between the ‘formal’ and the ‘informal’. In regard to the respect of commitments and the choice of non-opportunistic behaviour (e.g., paying off a sum due), individual tradeoffs depend on the credibility of the enforcement, the benefits and rewards.\(^{30}\)

### 5.3.3 The Criterion of Enforcement

Rules and institutions are cognitive representations that individuals find relevant—more credible, less cognitively costly in a given environment. These cognitive representations have the property of being normative, that is, of providing instructions for other representations and behaviour.\(^{31}\) Effective enforcement is an essential dimension of institutions as rules of the game: for individuals rules must constitute binding cognitive states (Aoki 2001). A key dimension of all institutions, formal and informal, is that they are individually shared beliefs, which bind behaviour at a collective level. Both informal and formal institutions get their binding dimension from their links to other institutions (e.g., an informal credit transaction from the common kinship of the contractors) and are subjected to historical change.
Power relationships also shape the particular actualizations and transformations of institutions, both formal and informal, for example the capacity for an individual to exercise her rights (such as land rights). Informal and unwritten rules are as coercive, enforceable and subjected to sanctions as are ‘formal’ laws and judgements. Conversely, North’s formal regulations (constitutions, laws, property rights) may be ignored, especially in developing countries with weakly institutionalized states. Group norms are coercive because they represent low-cost trust in uncertain environments and high costs (exclusion) in the event of a breach; they also represent information transmission and coordination of punishment supporting a reputation mechanism and in a self-fulfilling way (economic losses in case of non-compliance) (Akerlof 1976; Greif 1993).

Informal contract enforcement mechanisms may be even more credible on account of motives such as preserving personal relationships and trust, and formal enforcement (courts) may be viewed as less relevant. Informal contract enforcement may similarly be more efficient: informal lenders may enjoy higher loan repayment rates than formal lenders without litigation in courts (Aryeetey 1998). The presence of states is not pertinent in order for these mechanisms of enforceability to function. The various mechanisms underlying the credibility and compliance with rules have more explanatory power as to the effectiveness of rules than the dichotomy formality vs. informality.
5.3.4 Form and Contents of Institutions

‘Formal’ institutions and norms are often analysed as antagonistic vis-à-vis ‘informal’ ones: formal norms may erase informal norms or be ‘captured’ by them, which is viewed as a factor in the failure of reform in transition and developing countries (Hoff and Stiglitz 2004). There are, however, no theoretical grounds for such an intrinsic difference of nature between both categories. Similarly, the form of an institution is often simply viewed as being the totality of the institution in question, for example the legal protection of property rights offering effective protection. Yet all institutions include both an organization (‘form’) and implicit, traditional, path-dependent contents and meanings. All institutions are characterized by their ‘forms’—their definitions, names, modes of organization—and ‘contents’—their meanings, functions, relevance, and elements. These two levels are subject to distinct and continuous transformation due to historical processes and their environments. Institutional forms and contents are resilient or disappear depending on their relevance and credibility in the minds of individuals, which in turn depends on their perception of the events, history and the environment to which they are exposed.

Institutional forms have particular contents, relevance and functions on account of (a) their combination (e.g., hierarchical or complementary)34 with other institutional forms and contents and their exposure to the broader environment and (b) the absence of certain institutions. For example, in the context of weak state institutions rural institutional forms may be endowed with the functions of the state, such as managing taxation, redistribution, and equity issues, while the forms of public institutions may be perceived as individual predatory activities. The effective content of economic institutions may depend on the ways in which they are articulated with other
institutions, especially political institutions. This link determines, for example, the efficiency of a revenue service, which may otherwise be an institutional form that is ‘filled’ by various rationales (personal interest, creating employment, fulfilling solidarity obligations, among other things). Similarly, the functions and credibility of a political institutional form such as courts may depend on an executive and a legal system that are also credible and accountable. On the other hand, a particular function may correspond to various forms of institutions. ‘Informal’ institutions are no different, as they also have evolving forms and contents. For example, individuals may assign a market content to kinship reciprocity norms—they may use these norms, for example, as assets in labour markets (as in the case of the exploitation of kin members in a farm or enterprise).

Institutions are not defined ex ante by specific, necessary and sufficient sets of forms and functions: in the course of time and institutional transformation, certain combinations are perceived by individuals as being more credible than others and therefore may be enforced. Functions and meanings may become irrelevant, whereas the institutional form remains intact (e.g., traditional systems of reciprocity). The specific interpretation of institutions by individuals, the ways they shape individual behaviour and their modes of transformation are ex ante under-determined for ‘formal’ as well as ‘informal’ institutions.

5.4 CONCLUSION

Some dichotomies give rise to as many problems as the solutions they are aimed to provide. This chapter has shown that the concept of informality may not improve the
understanding of the phenomena to which they have been applied. This dichotomy
gathers empirical facts that are heterogeneous and has different meanings depending on
the theoretical approach and method. Furthermore, characteristics of informality that
could be specific to it are difficult to find. A distinction between institutional forms and
contents has been suggested, as has been a focus on criteria such as credibility and
enforceability. They may more accurately explain the dynamics of informal activities
and institutional transformation than does the formal-informal dualism.

REFERENCES


ENDNOTES

1 The author thanks the participants in the EGDI-WIDER conference for their comments, and particularly Ravi Kanbur for very stimulating suggestions, though the usual caveat applies.

2 ILO (1972); a history of the concept is in Bangasser (2000).

3 The fuzziness of the concept of informality is highlighted since the 1970s, see Bromley (1978), Peattie (1987).


5 Lewis (1954); the model is also presented in Bardhan and Udry (1999).

6 Bourguignon and Morrisson (1998) show that it has been assimilated with other dualisms, such as rural-urban income differentials.

7 Morrisson and Mead (1996) show that there is little relationship between different measurements based on size, registration, payment of taxes, and conformity to labour regulations.

8 See Johnson et al. (1997) for an estimation via the electricity consumption.

9 See the OECD Handbook (2002).

10 For data, see ILO (2002a; 2002b).

11 15th International Conference of Labour Statistics.

13 But not for the labour statisticians of the 17th ICLS (2003) who maintained the separation between the informal sector and informal employment; see Hussmanns (2004b).

14 Within the 1993 System of National Accounts (SNA93), see OECD et al. (2002).

15 And may be ‘national tales’, see Tanzi (1999).

16 As argued by De Soto (1989).

17 Highlighted by Hart (1973: 78).

18 Including civil service jobs, see Sindzingre (1998).

19 In Morrisson (1995), two-thirds of microenterprises were paying taxes in Thailand and three-quarters in Tunisia.


22 For example than formal African firms, as in Kenya, see Bigsten et al. (2000).


24 Such as the ‘bazaar economy’, defined by Geertz (1978) as repeated exchanges in a ‘noisy’ and uncertain environment.

25 This has been investigated by cognitive science and anthropology, see Goody (1986), but rarely by development economics.
26 Examples of complex traditional kinship, political and ritual systems being innumerable.

27 North (1996) on cognitive science, Denzau and North (1994) on mental models and institutions.

28 On credibility as a key issue in sub-Saharan Africa, Collier and Patillo (2000).


31 This cognitive theory of relevance is developed in Sperber and Wilson (1986); see also Sindzingre (2003).

32 Fafchamps (1996) on the example of commercial contracts in Ghana.

33 As shown by Greif (1997).

34 On complementarity and hierarchy as explaining the transformation of institutions, Aoki (2001).

35 Market institutions may also be ‘filled’ by kinship norms, as shown by Hoff and Sen (2005).

36 As highlighted by Rodrik (2002).