Legal capacity and the theory of the firm

Simon Deakin *


1. Introduction

The idea of legal personality entails, more precisely, the attribution of capacity, that is, the ability to hold property, make contracts, and carry out other acts with legal significance. Capacity is the entry point for participation in a market economy in which economic relations are legally constituted. For private law to attribute capacity to all human persons was not a natural or inevitable consequence of legal development but part of the wider political reconfiguration which brought about modern constitutionalism and the recognition of universal human rights. Attributing capacity to private organisations was not a natural step either, and the contemporary evolution of corporate law suggests, in various ways, regression to pre-modern forms of social order.

To explore this idea, this paper will first of all provide an overview of a theory of juridical concepts, which refers to their nature (‘ontology’) as evolving linguistic forms (section 2). This is followed by a functional account of the legal concept of capacity as the gateway to participation in a market economy (section 3). Section 4 contrasts the notions of ‘personal’ (human) and ‘corporate’ capacity. Section 5 concludes.

2. The ontology of juridical concepts

A juridical concept can be defined as a linguistic form used for specifying the conditions for the application of a legal rule which takes the form ‘in C, if X than Y’, as in, ‘in the context of a company limited by share capital, if X is an director, he owes the company fiduciary duties of loyalty and care’. Several concepts are embedded in such as rule: the terms ‘company’, ‘director’; and ‘fiduciary’ all have to be unpacked.

It might be thought that the concepts themselves add relatively little to the content of the rule. The concept ‘fiduciary’ could just be shorthand for stipulating the type of position which gives rise to obligations of loyalty and care. The concept is an intermediating link or bridge between a set of factual preconditions, on the one hand, and certain normative or deontic conclusions, on the other. It can be said to represent or condense a good deal of otherwise unwieldy information, but without adding anything of its own. This was Alf Ross’s view: the concept is ‘inserted between the conditioning facts and the conditioned consequences’ solely as ‘a means of presentation’. It is ‘in reality a meaningless word without any semantic reference whatsoever’ (Ross, 1957).

Ross’s point was that, after allowing for the representational effects of concepts, we should not ascribe independent causal powers to them. In particular, we should be very cautious to accept that concepts, in themselves, determine the content of rules or the conditions for their

* CBR, University of Cambridge (s.deakin@cbr.cam.ac.uk). This paper represents work in progress. It builds on my join work with Alain Supiot and colleagues on the Capacitas project (Deakin and Supiot, 2009) and on my contribution to the work of the Cambridge Social Ontology Group (Deakin, 2014).
application. Such caution is undoubtedly justified. However, as Giovanni Sartor has argued, Ross’s view is excessively reductive. This is for a number of reasons.

Firstly, concepts acquire inferential meaning from the prior conditions for and subsequent effects of their application, but also create meaning by virtue of the role they play in integrating empirical and normative data. Thus the definition of the prior conditions for the application of a rule can never be entirely separate from the consequences of applying it: there is ‘a feedback circle involved in construing legal norms, connecting assignment of a meaning to a term and a teleological interpretation of the norms including that term’ (Sartor, 2009: 27). Thus the normative context in which a rule is applied is unavoidably taken into account when the prior conditions for its application are being considered.

Secondly, concepts frame the possibilities of legal interpretation by virtue of the taxonomical order in which they are arranged (Sartor, 2009: 19). Concepts are linked to each other through subdivision and fragmentation, with subcategories and satellites expressing exceptions from or variations of originating or ‘foundational’ terms. Conceptual meanings are derived at least in part from the positions of particular terms within the overall taxonomical order.

Thirdly, concepts assist the process of legal interpretation through their open-endedness or ‘defeasibility’ (Sartor, 2009: 28). Concepts are useful tools because of their incompleteness and openness to new applications. Concepts can shift their meaning over time; their form, as well as their content, can evolve by reference to changes in society and the economy. Because they serve as a bridge between facts and norms, legal concepts are continuously open to empirical data.

At the same time, the reception into the legal system of data from the social realm beyond the legal text involves a process of translation. This insight, which is at the core of the systemic view of law associated with Niklas Luhmann and Gunther Teubner (Luhmann, 1984, 1993; Teubner, 1989) is often understood as generating a theory of ‘closed systems’, which is indeed one of the aspects most stressed by its authors. However, autopoiesis, or self-reference, is only one part of the theory; the other is the capacity of systems to evolve by reference to their environment. The legal system is simultaneously closed and open, that is closed at the level of its internal operation, which consists of recursive self-reference, but open at the level of its co-evolution with and adjustment to other social sub-systems.

The ontology of systems theory is both naturalistic and constructionist (on these categories, see Searle, 1995, 2010; Lawson, 2014). It is naturalistic in so far as it conceives of the social realm as a context governed by ‘laws’ in the descriptive sense, that is, regularities, which are consistent with the laws governing living forms in the natural realm. The social realm is a subset of the natural one; there can be no society, no realm defined by human interaction, without the existence of human beings who are, first and foremost, physical entities, that is to say, psychic or biological systems in their own right. The ‘laws of form’ which determine the structure of living systems have their equivalent in regularities which frame the structure of society. The task of the human and social sciences, in identifying and unravelling these regularities, is no more or less ‘scientific’ than the project of natural science disciplines such as molecular biology or population genetics. In both cases, there is an external ‘world’ to be mapped and analysed using relevant techniques of inquiry.
However, the social realm is not reducible to the natural one. The working hypothesis of systems theory is that the societal ‘laws of form’ bear a family resemblance to, but are not identical with, those found in the natural realm. Biological concepts, such as self-reference, inheritance, mutation, and so on, are useful starting points for the theory of social systems, but are not unaffected by the process of their application to societal data. This is perhaps the problem with the idea of the ‘meme’, which has been proposed as a fundamental unit of cultural selection by analogy with the gene in natural selection. There is nothing wrong in principle with the idea of drawing analogies from the natural sciences for the purposes of theorising the human and social ones, but proponents of the meme are looking in the wrong place when they associate the meme with psychic or neurological systems. If genes have their equivalents in the cultural sphere, they are to be found at the level of phenomena such as conventions, institutions and norms which express elements of a distinctively social structure (Deakin, 2003).

According to the theory of social systems, if the natural realm consists of ‘matter’ in the form of quarks, atoms, genes, proteins, and so on, the social realm consists of ‘communication’. To put it this way seems deeply counter-intuitive; does society not consist at least of individuals and possibly, through their mutual interaction, of enduring institutions? Of course, ‘communication’ cannot exist without human beings to do it, and it is their behaviour which constitutes ‘communicative action’. However, the two sets ‘communication’ and ‘behaviour’ are not identical. There are aspects of human behaviour which are rooted in physical, for example genetic causes, and there are aspects which are social in origin. It is the study of the specifically communicative dimension of human behaviour with which the social sciences are concerned. This is an important methodological move as it helps the social sciences to avoid the excessively reductive strategies of fields such as ‘sociobiology’ and ‘neuroeconomics’, which attempt to offer all encompassing naturalistic explanations for societal phenomena. Systems theory can perfectly well accommodate the existence of a natural, biological substrate to society, while proceeding on the basis that social structures have emerged out of, and hence are distinct from, that material base.

Conversely, communication can take non-behavioural forms. Data can be communicated through human speech and through non-verbal behaviour, but they can be also be embedded in physical objects (such as a traffic light or a banknote) and in texts (such as legal judgments and statutes). Complex texts ‘script-code’ behavioural regularities (Deakin and Carvalho, 2011). These texts are not free-standing; they would not exist were there no human beings to make them, whether directly or at one remove through technologies of various kinds. However, organised texts, such as those of the legal system, display features which are not reducible to the psychic or behavioural strategies of the human beings who are their authors. In particular, texts have their own laws of form, that is, their own evolutionary tendencies. This entails a second important methodological move: systems theory invites us to study both behaviour and text, ‘the world’ and ‘the word’, and the links between them, in an evolutionary, that is to say, historical, perspective.

It follows that the social reality assumed by systems theory is a ‘constructed’ one: it is the result of multiple discursive practices, each of which is self-reproducing, and which are constantly coevolving by reference to one another. The view of the social world constructed through legal discourse is distinct from the perspective of the economic or political system. A term such as ‘company’ (or ‘corporation’) acquires its legal meaning from the place of contract within the taxonomical structures of juridical thought, which are unique to legal reasoning. There is no precise match with the same term in economic discourse or in
everyday language. Furthermore, there is no single, overriding social category of ‘company’ to which the different discursive practices of the separate social systems have to conform or from which they can be said to originate. There are only multiple, overlapping and continuously mutating discursive forms.

It does not follow from the communicative nature of social reality that there is no external world beyond the text. The text is the social world, but, importantly, only one part of it. Thus the study of linguistic structures, their presuppositions and commitments is one aspect, but only one, of social science method. Communicative data is embedded in texts, but also in routines, conventions and norms which operate at the intersection of language and behaviour, and in physical objects which perform a communicative function.

Nor does the ‘closure’ of the various social sub-systems carry the implication that there are no points of correspondence between them. Legal concepts are separated from other dimensions of the social realm by the autopoietic closure of the legal system at the point of its self-reproduction. Legal concepts refer to other concepts, legal rules to other rules, and so on. This is a feature of law’s autonomy, of the neutral ‘rule of law’ which is a core feature of modern liberal-democratic polities. But this ‘operational closure’ is counterbalanced by the system’s ‘cognitive openness’, that is, its capacity to absorb signals from its external environment and to translate them into the terms of its own operation. Put more concretely, concepts in the legal system will be more or less successfully aligned with their referents in the economic realm, for example, to the extent that they can translate the transactional logic of economic relations into terms present in juridical language.

It is this tension between openness and closure which drives legal evolution. This process can be understood through the operation of a ‘variation, selection retention’ algorithm which has points of correspondence with the same process in the evolution of living systems. On the one hand, conceptual categories, expressed through taxonomies which are often quite rigid and unbending, provide the skeleton frame or infrastructure which prevents the mass of individuated norms dissolving into chaos. In evolutionary terms, they provide the basis for the ‘retention’ or ‘inheritance’ of forms which has its biological equivalent in the inter-generational transmission of genetic structures. On the other, concepts provide the gateway for empirical data to enter the legal system. They condense or code information about the social realm beyond the text, in such a way that makes it possible for these empirical data to influence the content of legal rules. In this way, mutation or variation of legal rules is possible. Selection of rules, reflecting pressures from the external environment for rules which are more or less functional in given social settings, is internalised through the same distinct forms of juridical reasoning: legal rules persist not simply because they ‘fit’ with external environment parameters, but according to how far they operate consistently with the internal categories of legal analysis.

It follows that we should not expect legal evolution to produce rules which are optimal for a given set of external economic or political conditions. The law cannot be completely open to its environment, since that would imply the dissolution of the system itself. While the preservation of a self-reproducing system might not seem to offer any advantage to society in terms of its broader functionality or fitness, the point is that a social world without such boundaries would be informationally impoverished. The autonomy of the legal system is the functional precondition for the preservation of the applied information (‘knowledge’) embedded in it. This information is useful, possibly essential, for achieving social coordination in an otherwise endlessly uncertain and contingent world.
But it is also clear that the legal system can be expected both to respond at some level to changes in its context, and to initiate them. The law, then, is partially endogenous to economic and political change. It can be both the independent and the dependent variable. On the one hand, regularities which begin in a transactional or organisational context can find their way into the legal system. Conventions which are based on shared knowledge among a population of actors can emerge on the basis of repeated interactions between them (Lewis, 1969). The law may adopt the content of conventions: ‘general clauses’ in contract law provide a portal for commercial practice to enter legal discourse (Teubner, 1989).

3. The evolution of capacity

The concept of capacity has undergone a series of evolutionary transitions which reflect the changes which have taken place in industrial societies since the rise of the market economy in the late eighteenth century, a process which coincided with the formalization of modern private law in the civilian codes of continental Europe and the more or less systematized common law of England and North America. Studying changes in the idea of capacity over time is one way to throw light on the underlying structure of a market economy and of the place of law within it (Deakin, 2009).

The juridical idea of capacity signifies a status conferred upon persons for the purpose of enabling them to participate in the economic life of the polity. In modern legal systems, ‘capacity’ is the principal juridical mechanism by which individuals and entities are empowered to enter into legally binding agreements and, more generally, to arrange their affairs using the instruments of private law. Capacity, in the legal sense, is thereby the gateway to involvement in the operations of a market economy. In its classical form, originating in the contract law of the nineteenth century, capacity was defined negatively, that is to say, by its absence: the very young, very old or very ill were deemed, to varying degrees according to particular contexts, to lack the ability to make legally enforceable contracts. The concept of capacity was therefore a doctrine of selective contract enforcement, which both protected the incapable from exploitation, but, equally importantly, protected the market against the incapable.

Alain Wijfells (2009) has traced the historical development of the concept of capacity and related its emergence, in broad terms, to the rise of a liberal market order based on the principle of freedom of contract, or respect for the autonomy of the contracting parties. Legal classification matters, as he reminds us, because changes in the categories of legal thought provide an insight into the wider forces at work in a given period. The rise of the idea of capacity in the late eighteenth and early nineteenth centuries was part of a process Wijfells refers to as ‘rationalisation’, meaning the displacement of heterogeneous legal categories, typical of the hierarchical and status-based organization of society under the Ancien Régime, with relatively homogeneous ones, based on the notion of the equal status of all citizens: ‘every single human being – and only the single human being – enjoys capacity’, as Savigny put it. From this was derived the legal presumption that all physical persons had full capacity, with the remaining categories of incapacity (relating principally to minors, married women, and the physically and psychologically incapable) viewed not as particular types of status but as exceptions to the general rule. These exceptions were increasingly narrowly expressed over time.
Notwithstanding its orientation in favour of freedom of contract, the classical core of contract law gave expression to a theory of the institutional preconditions of a market economy, albeit a rather minimalist one (Deakin, 2009). Incapacity, as a ground for invalidating contractual agreements independently of misrepresentation, fraud or force, was premised on the idea that certain persons did not possess the power to make rational assessments of their own self-interest of the kind required for market-based exchange. The contract law of the nineteenth century also recognized that certain types of transaction could be denied contractual force on the grounds that they infringed particular values which, exceptionally, took priority over the value of freedom of contract. In the common law, these went under the heading of ‘public policy’, and their effect was to qualify the power to make binding agreements which was otherwise generally vested in physical persons. This part of private law therefore offers fragments of a theory of functional limits to freedom of contract: these limits are necessary both to preserve the market against anti-competitive behaviour (as in the case, for example, of the doctrine of ‘restraint of trade’), and also to preserve society itself against the market (for example, the rules against the enforcement of certain illegal or oppressive contracts).

These were no more than fragments, though; the common law notion of public policy was and is an extremely limited one which was frozen in time at the end of the nineteenth century. Social or regulatory legislation has since become a far more significant source of contractual regulation. Wijfells (2009) argues that in the rise of regulatory legislation we can observe a process of ‘derationalisation’, which has led to the fragmentation of the unitary concept of capacity: new forms of legal protection have arisen, representing new interests. In areas such as consumer or employment protection, the law has qualified the notion of capacity, without using the technique of incapacitation to do so. The unitary category of capacity is being eroded by a set of discriminating and differentiating techniques. This poses the question of whether the abstract principle of equality can be maintained in the face of the proliferation of interests claiming protection.

4. Comparing personal capacity and corporate capacity

It is not the case that the attribution of capacity to entities involved the translation of an existing concept, legal personality, which had first been developed for human beings, into the context of organisational forms. Certain entities, including religious associations, guilds and governmental bodies, as well as trading corporations supported by the state, acquired legal capacity several centuries before the point at which it came to be seen as an inherent right of human persons. The law of the Ancien Régime was based on multiple forms of personal status, which were intended to apply in particular contexts. The private law codes of the nineteenth century replaced a polycentric and fragmented legal structure with a single, centralised juridical regime, which was intended to apply common legal principles to the full range of social and economic relations. The emergence of the meta-concept of the ‘legal person’, capable of applying equally to human beings and organisational entities, was part of this process. It did not signify in any sense (occasional legal metaphors aside) the attribution of the physical or biological characteristics of human agents to corporate entities. Instead, it indicated the extension, to individuals and organisations alike, of the legal-economic regime of private law and the market order.

In relation to individuals, the concept of capacity formalised the idea that participation in market exchange presupposed the ability of agents to take an informed and consistent view of their own interests. In this sense, the idea of capacity in private law is the juridical equivalent
of the concept of rationality in economic theory (Deakin, 2009). Both ideas have their origins in nineteenth-century liberal thought.

In civil law systems of mainland Europe and those parts of the world influenced by civilian legal thought, contractual capacity is one of the basic conditions for the formation of a contract. Martijn Hesselink (2009) shows that, in the civil law, ‘capacity’ has two meanings, one referring to the ability to hold rights (capacité de jouissance, Rechtsfähigkeit, capacité juridical and so on), the other (capacité d’exercice, Handlungsfähigkeit, capacità di agire, and so on) the ability to exercise them. Contractual incapacity is almost invariably an instance of second of these two categories. The capacity to hold rights is in effect a right to be treated as a legal subject and not as a res, that is, an object of legal relations. That this right should presumptively vest in all human persons from the point of birth, and over time has come to be recognized not just as a foundational principle of private law but as an aspect of human rights. This is equally the position in the common law, where the formal distinction between the holding and exercising of rights is, however, of no relevance. In all private law systems, grounds of incapacity are now tightly defined; they mostly apply to agreements made by ‘minors’ (or ‘infants’), the very old, and those suffering from mental illness. The question of contractual capacity has been said to be of ‘reduced practical significance’ in the English law of contract; although not without some theoretical interest, particularly from the point of view of the interaction of contractual and restitutionary remedies, it is essentially treated as a footnote to the main body of the law of contract. The relative insignificance of the subject for English lawyers is compounded by the absence of certain complex rules of French-origin civil law systems, dealing with the circumstances under which lack of contractual capacity can be offset by assistance (curatelle) or representation (tutelle). These have no equivalent in the English common law, and even some civil law systems, such as the German one, do not recognize the category of assistance. In all systems, the effects of incapacity on a contract also differ according to context; complete nullity is only one option, others including voidability, while alternative remedies in tort or restitution may be available.

When we view company law through the lens of capacity, we can straight away see that much of it is concerned with the same issues of representation which arise in the context of the incapacity of human beings. Companies can only act through human beings, and so company law tells us which individuals, and under what circumstances, have the power to bind the corporate entity. Company law also sets out limits on the capacity of the company to undertake particular types of commercial transaction, by reference to the terms of the contract made between the company’s members, through what English law refers to as its memorandum of association. Over time, the tendency of the law is to allow a larger range of corporate officers to bind the company through their own acts, while removing constraints on the capacity of the entity which are associated with the ultra vires rule. Both tendencies reflect the view that third parties transacting in good faith should be protected as far as possible from the consequences of the absence of agency, on the one hand, and narrowly defined corporate powers, on the other, in terms of the nullity or voidability of contracts.

Granting capacity to entities enables them to access the modalities of market-based exchange in just the same way as it does for individuals. ‘Asset partitioning’ protects the assets of the company from third party claims against its members, while also entailing the exposure of the same corporate assets to claims incurred against the enterprise as a continuing organisation. Credible contracting and ‘bonding’ is facilitated by the permanence of the corporation as a legal form, that is, by its depersonalization. In all these respects, legal personality has
material effects, although these should not be exaggerated: adaptations of the concept of the trust enabled manufacturing firms in England to raise equity capital without the need for incorporation during the period of the industrial revolution when very few such enterprises had corporate status.

Granting capacity to entities has wider effects: it integrates firms and associations into the market economy, in just the same way as making capacity generally available to human persons entails their participation in the market order. Through the device of legal personality, entities can hold property and become counterparties to contracts. In so doing, they open expose to new forms of liability for debts and harms. The idea of corporate responsibility has a specific juridical root: long before it became conventional to refer to CSR in terms of the voluntary assumption of enterprise, responsibility, the law had established the liability of the enterprise to its contractual counterparties and to third parties exposed to harms arising from its activities.

The transformation of the idea of personal (human) capacity which has occurred with the rise of the regulatory state has also affected the concept of corporate capacity. The advent of the welfare state saw the emergence of new techniques for countering the risk of exploitation in highly unbalanced or unequal contracts – as Jean Hauser (2009) puts it, a shift from ‘incapacitation’ to ‘protection’. Where the classical, nineteenth-century private law concept of capacity provided protection to the weak or vulnerable by denying legal enforcement to their contracts, thereby excluding them from independent participation in economic life, statutory regulation of the twentieth century inserted mandatory and default terms into contracts for the benefit of parties deemed to be at a disadvantage in terms of bargaining power.

Sandrine Godelain (2009) examines the evolution from incapacitation to protection in consumer protection law and labour law in France. Once the civil law ceased to see the technique of incapacitation as the main solution to extreme inequalities of resources and information in contracting, it adopted a number of alternative techniques for dealing with highly unequal contracts; in the consumer law field, the doctrines of mistake, fraud and misrepresentation were all deployed to deal with the problems posed by contracts giving rise to excessive indebtedness. However, these interventions were mostly ineffective. A more productive route emerged when lenders were placed under a duty to make an informed assessment of the ability of a consumer to take on large amounts of debt; hence the focus on the financial capacity of the borrower. In labour law, the idea of the professional capacity of the worker, understood as his or her ability to participate in a labour market increasingly characterized by flexibility, multi-tasking and career shifts, has been coalescing around a number of developments in the laws governing the employer’s duty to provide training and the state’s involvement in employment policy.

What does the rise of the regulatory state signify for the nature of corporate capacity? The corporate form now appears not just as the object of regulation in areas such as consumer, worker and environmental protection, but as an actor in the regulatory process. The corporation is being used to fulfill a variety of regulatory goals, such as loss spreading and risk diffusion. The boundaries of the firm become coterminous, for legal purposes, with the limits of enterprise liability, as in the tort law concept of ‘enterprise risk’ (Deakin 2002). There are numerous juridical techniques for identifying the corporation’s asset pool which including the lifting of the corporate veil, the imposition of joint and several liability on firms
in supply chains, and the attribution of group identity to companies in a parent-subsidiary relationship.

Regulatory avoidance is the converse of the process: the use of the corporate form to partition assets in a way which deflects the application of regulations. We see this most clearly in tax avoidance, where it intersects with jurisdictional competition to enable enterprises not just to escape existing standards but to instigate a race to the bottom. This is not at all a new problem; in the common law, it goes back to the origins of the corporate form in the use of trust, which was invented precisely in order to enable property to be protected from the fiscal and regulatory powers of the medieval Crown. The trust has of course long since evolved to acquire many other functions, but its original purpose continues to influence the development of the law, and to limit statutory and judicial attempts to counter tax avoidance.

Once legal capacity was attributed to corporate entities and not just to the ‘single human person’, the equality inherent in the classical private law concept of capacity was undermined, and the conditions set for the concentration of wealth and power through the corporate form which has become a feature of contemporary capitalism. This tendency is taking us back to a fragmented world, similar to the pre-modern social order, characterised by hierarchical forms of status. Unpicking this process is not straightforward. It cannot be done by a strategy of incapacitation, that is, by denying legal validity to acts of the enterprise. Techniques of identification, tying the corporate form to the asset pool of the enterprise, offer a more plausible response.

5. Conclusion

The juridical idea of capacity is a fundamental legal concept which acts as a bridge between legal discourse and social relations of the kind which characterise a market-based economic order. Granting full capacity to corporations as legal persons in their own right had various material consequences for the organisation of capital, but also, more generally, had the wider effect of extending the logic of the market order to entities which had previously been partially insulated from it. This is the parallel process to the one which occurred in relation to human beings once it became accepted that ‘every human person has capacity’. Participation in a market order implies not simply freedom to hold property and to contract, but also legal responsibility for debts and harms. The idea of corporate responsibility has a legal root which is in danger of being forgotten in the contemporary emphasis on the voluntary nature of CSR. In modern societies the corporate person is not just an object of regulation but a regulatory actor in its own right, increasingly used to deliver public policy goals. By contrast, the use of the corporate form to evade taxation and regulation owes more to pre-modern notions of a stratified and fragmented legal order. To permit the corporate form to be used for regulatory avoidance on the scale which has become commonplace today is to revive the legal thinking of the Ancien Régime.

References


