Journal of Institutional Economics (2022), **18**, 361–378
doi:10.1017/S174413742100059X

CAMBRIDGE
UNIVERSITY PRESS

RESEARCH ARTICLE

Mental models and institutional inertia

Eckehard Rosenbaum

European Commission, Joint Research Centre, Via Enrico Fermi 2749, 21027 Ispra, Italy
Corresponding author. Email: eckehard.rosenbaum@ec.europa.eu

(Received 1 December 2020; revised 2 July 2021; accepted 5 July 2021; first published online 3 August 2021)

Abstract

Institutional inertia as one of the underlying reasons for hysteresis is often ascribed to external factors such as the distribution of wealth and income. Complementing these findings, the paper focuses on important internal factors, which render institutions stable and which prevent fast institutional changes, namely the role of mental models. Their importance is derived from the analysis of an important set of institutions, which can be described as enabling rules. Such rules enable actors to do certain things, such as speaking a language or playing chess. In doing so, enabling rules arguably require complementary mental models, which contain not only knowledge about the rules and the context in which they are applied, but also about how to apply the rules successfully. An important implication of this conceptualisation is that institutions and their representation are interdependent and mutually stabilising.

Key words: Hysteresis and inertia; institutional change; mental models

1. Introduction

Structural reform is, and has always been, an important aspect of government policy. While the notion of structural reforms is sufficiently unspecific to accommodate a broad range of policy measures with varying objectives and sometimes even contradictory measures, structural reforms can nevertheless be seen as changes to the framework that determines how government, society and economy work. Thus, what is supposed to be reformed are the socially shared rules of the game (Dequech, 2013), i.e. the institutions that shape and – to some extent – govern political, social and economic processes.

Against this background, Palley (2017) has argued that such policies are also liable to affect the politico-institutional order and thereby the conditions under which policies are formulated and decided upon in the first place. For instance, institutional changes may affect the distribution of wealth and income and thus of the very endowments which give agents the means to prevent certain policies (Palley, 2017). Consequently, institutions exhibit *inertia*: they themselves contribute to strengthening the conditions that render intentional institutional change more arduous. Without downplaying the importance of the various mechanisms identified by Palley (2017), they mostly concern what can be termed *external* constraints on institutional change (such as a specific distribution of wealth and income). In other words, the distribution of income and wealth, to continue this example, has an impact on institutional change, whether or not the institution in question also has an impact on that distribution.

The point to be developed here is therefore that there are arguably also *internal* constraints resulting from the very nature of institutions as key ingredients for socially reproduced structures and the role of mental models therein. This suggests that further insights can be gained by complementing the investigation of inertia by a broader account of institutions and institutional change. Such an endeavour appears all the more worthwhile as a wide spectrum of economists and social scientists has recognised the importance of institutions not only for economic performance but indeed for creating the very fabric that nits our societies together and, ultimately, constitutes them (Hodgson, 1988; North, 2006; Searle, 1995).

As I shall argue, many institutions are stable because of the way they are represented in mental models and reproduced through social interaction shaped by such models. I shall further surmise that institutional inertia resulting from such representations is Janus-faced insofar as it is both, a precondition for institutions to function, and a factor, which prevents or slows down attempts at achieving institutional change. Beyond identifying another reason for institutional inertia, the paper therefore seeks to contribute to the theory of institutions and institutional change. In doing so, it suggests that only a deeper understanding of social reality as advanced by Tony Lawson and others (Lawson, 1994, 2003, 2012a, 2016a, 2016b) and informed by psychological reasoning can form the basis for a further development of institutional theory.

It would go way beyond the scope of this paper to provide a comprehensive account of institutions and institutional change. What I intend to do is to initially focus on two aspects of institutions, which have hitherto received perhaps less attention than is warranted but which have, or so I shall argue, implications for comprehending institutional inertia. The first aspect concerns different understandings of institutions. In particular, I shall argue that the widespread understanding of institutions *as constraining rules* downplays their equally important role as coordinating and enabling devices. Consequently (this is the second aspect), it has been criticised with arguments, which are not fully convincing given the variety and function of institutions and the various reasons for complying with these institutions. In doing so, I hope to strengthen an account of institutions as rules while adding some aspects to the debate on conformity with institutions.

These aspects then lay the groundwork, or so I wish to show, for the development of a two-dimensional typology of rules. Based on this typology, I shall then argue that many institutions *as socially reproduced and enabling structures* require for their procreation through intentional action not only to be mentally represented (Denzau and North, 1994), but such mental models must in a way be seen as important bearers of institutions, epitomising as they do the very rationale of, and for, institutions.

The paper is structured as follows: section 2 discusses different types of institutions while section 3 argues that the notion of mental model as suggested by Denzau and North (1994) or Jones *et al.* (2011) is a necessary complement of, in particular, enabling rules. The second part of the section discusses a number of critical arguments that have been made against mental models and the theoretical framework used here. Section 4 addresses, albeit briefly, some empirical work and its policy implications. Section 5 concludes.

2. Institutions as/and rules

Towards a typology of rules/institutions

Hodgson (2006) defines institutions as systems of established and prevalent social rules that structure social interactions. A rule is then 'broadly understood as a socially transmitted and customary normative injunction or immanently normative disposition, that in circumstances X do Y' (Hodgson, 2006) where '"do" is to be interpreted as a placeholder for phrases such as "this counts as", "take this to mean", "refrain from" and so on' (Faulkner and Runde, 2013, following Lawson, 2012b). These are the notions I will employ in what follows. Accordingly, I take it that language, money, law,[1] systems of weights and measures, table manners and firms (and other organisations) are all institutions in this sense, and so are of course property rights. The above list suggests that the content and form of

[1] Fleetwood excludes laws and regulations on the grounds that they are properties of organisations, not institutions (Fleetwood, 2008). However, while the laws governing organisations may not be institutions of the kind, which is of interest here, there are arguably many laws, which exist outside organisation and which differ from norms only because of their formalisation.

https://doi.org/10.1017/S174413742100059X Published online by Cambridge University Press

institutions may vary considerably. Some institutions are codified (money, law), i.e. encoded in formal rules, others are not (table manners), or only partially so (language). Some institutions are legally binding (mostly law), others are not (table manners). Importantly, institutions are not just patterns or regularities in the flux of events (Fleetwood, 2008), although they may give rise to such patterns or regularities. Any of the above-mentioned institutions continues to exist even in the absence of their concrete realisation.

In adopting such an understanding of institutions as rules, I do not wish to take a definitive position on a conceptualisation of institutions as the outcome of coordination games, as this would require a separate paper. Suffice it to say that there is one element of this literature that I think is relevant and that I will address. This is the notion that institutions as rules, beyond their constraining effects, also serve to coordinate behaviour. Institutions so conceived set out a framework within which human interaction takes place, including social positions and their associated rights and obligations. This is so because institutional rules do not only specify what can (not) or should (not) be done, but also by whom and to whom. Thus, institutions often assign functions and define roles insofar as institutions hardly apply to everyone in all circumstances, but only to some agents in some situations and at some times.[2] Institutions thereby enable the establishment and shaping of relationships between persons and, hence, an order, whose properties may transcend those of its constituent elements (Lawson, 2012a) and are in precisely this sense emergent. This order is what I consider to be a social structure. At any point in time, social structure is both inherited and in the making, it is *causal* – in the sense of making a difference (Lewis, 2005), and being reproduced – and in this sense *caused* – through action and interaction without this causation being the motivation for the action. In my understanding, this is also the essence of the transformational model (Lawson, 1994) of social interaction.

How is the term ‘social’ in the above definition is to be understood? Is it to be understood as mainly referring to an empirical categorisation of rule following or does it mean that a rule is followed by several individuals to avoid punishment, thus implying a causality? Guala (2015), for instance, has argued that ‘many social institutions do not rely on normative commitments engendered by a joint intention’ but that conformity with these institutions is ensured by threatening deviants with punishment. Indeed, the insistence on enforcement via punishment or another kind of negative incentive is often considered the essence of the institutions-as-rules approach. A key problem of this approach, or so its critics therefore argue, is that it treats enforcement as exogenous and thus must explain who enforces enforcement (Hindriks and Guala, 2015a).

Since I will largely follow the institutions-as-rules approach, some further remarks on the enforcement argument are in order. First, and without downplaying the importance of enforcement *for some rules*, it seems that many rules are indeed followed without being combined with positive or negative incentives other than the – culturally mediated – belief that these rules make sense (Hodgson, 2015b). Enforcement may in fact be a marginal issue in the sense that it matters only for those at the fringes of society. The others accept rules because they have been socialised to believe that these rules are intrinsically good for everybody, and these beliefs are in turn vindicated by the codification and possibly enforcement of such rules, but do not presuppose the latter.

At the same time, one may not only make a distinction between accepting a rule for fear of punishment or for moral reasons, but also for want of alternatives or due to the normative power of the factual (Dequech, 2013). Hence one could say that institutions do not need to be accepted or embraced; it is sufficient that they be recognised (Searle, 2015).[3] Finally, also epistemic legitimacy

[2] This is not to say that institutions assign functions to, or define roles for, *specific* persons like you and me. Rather, institutions determine how to select a person that is then to be given a specific function or that is then supposed to play a specific role (e.g. being President of the United States).

[3] While I do believe that Searle has important things to say with respects to the topics discussed in this paper, Lawson (2012a, 2012b) has highlighted that there are also important differences between Searle’s ontological conception and that endorsed by realist philosophers in the sense that the latter’s analysis as exemplified by Lawson (2012a, 2012b) starts from generalised features of human interaction and then works backwards from actual social interactions to their conditions of possibility. Searle, by contrast, seeks to investigate how anything that might be termed ‘human society’ has arisen out of

https://doi.org/10.1017/S174413742100059X Published online by Cambridge University Press

may contribute to conformity with (or resistance to change of) institutions (Dequech, 2013). Institutions are not questioned to the extent that they help us make sense of the world around us.

However, even if enforcement appears to be necessary for certain rules (or at least for certain actors), this is not a particularly strong argument against the institutions-as-rules approach. After all, while it is an interesting question why basic institutions such as certain state functions have emerged in the past, institutional development and change today do not take place in an institutional vacuum but against the background of a rich and varied social, cultural and institutional landscape. In such a setting, the issue of enforcement via both formal and informal means clearly looks different and does not lead as easily to an infinite regress. Enforcing institutions are already – and almost always – present and do not need to be assumed or explained. In particular, property rights do not exist in a vacuum but encompass acknowledged rights granted by legitimate legal authority and hence a state (Hodgson, 2015a).

Third-party enforcement also seems to be less of a problem for institutions that share an important characteristic, namely their ability to coordinate behaviour. While these institutions also constrain behaviour (if they wouldn't, they would be ineffective in coordinating), compliance results from the immediate consequences of individual actions 'out of equilibrium' (the angry responses of other drivers if not a crash) rather than the threat of later punishment and by a third party. But if this is so, then it is also preferable to maintain a rather wide and basically empirical understanding of the term 'social', which can accommodate various reasons for rule following. Such an understanding would suggest that rules not shared by at least several actors are not institutions in the above sense.

The foregoing observations on enforcement hint at an important feature and function of institutions though. As already suggested by Hodgson (2006), Searle (2005) and others, many institutions play an *enabling* role. That is, they allow us to do certain things in the first place which would not be possible, or at least much more difficult, without these rules. This peculiar and important role of institutions is not fully captured, or so it seems, by an account of institutions-as-rules, no matter whether the emphasis is placed on their constraining effects or their coordinating character. Enabling institutions both coordinate and constrain, but both their constraining nature and their coordinating role are (necessary but not sufficient) characteristics derived from the enabling function of these institutions.

To elaborate on this, let us look at various types of rules/institutions in more detail (see Figure 1). To do so I suggest that rules can be conceptualised in terms of two salient characteristics. The first is *restrictiveness*, i.e. the extent to which rules impose limitations on possible behaviour. The second important characteristic of rules is their *constructiveness*, i.e. the extent to which (sets of) rules allow us to do certain (hitherto impossible) things. Both features together help to characterise four generic types of rules. These four types capture, or so I would argue, most of the rules that make up institutions, and among those, above all enabling rules are closely intertwined with mental models in that enabling institutions are best seen as tools, which require both competence and knowledge for their use.

In proposing this typology, I do not wish to suggest an alternative to typologies based, for instance, on the degree of formalisation, viz. Douglas North's distinction between formal and informal rules (North, 1991) or Fleetwood's more recent discussion of the same matter (Fleetwood, 2019) – the latter will be taken up again in the section on mental models. Rather, the typology proposed here seeks to draw attention to a possible differentiation between rules that are analytically useful for understanding which purposes rules fulfil and what they presuppose or imply in terms of our cognitive understanding and knowledge.

material that is traditionally studied by the natural sciences. Thus, he appears to work forwards. Another difference highlighted by Lawson (2012a, 2012b) is Searle's insistence on the role of language as being prior to collective practices. By contrast, Lawson holds that 'language capable of representing rights and obligations, is in part built on, and presupposes, the (prior) existence of normative collective practices' (p. 364–5).

https://doi.org/10.1017/S174413742100059X Published online by Cambridge University Press

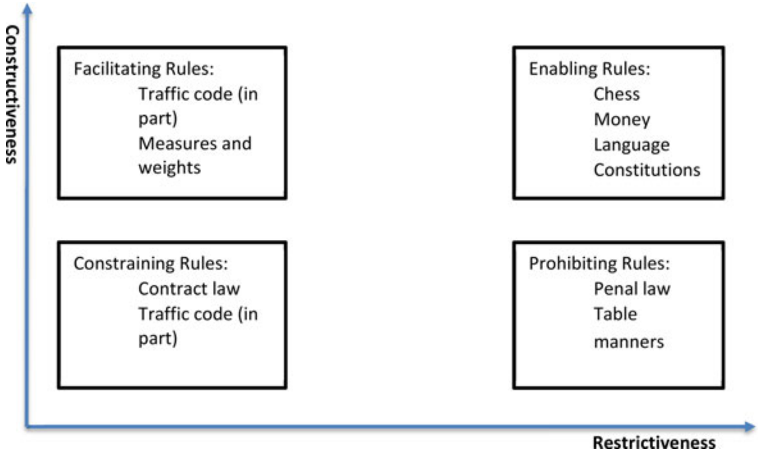

Figure 1. A typology of rules.

Prohibiting and constraining rules

'Thou shalt not kill', the Seventh Commandment, is probably one of the most well-known, and prevalent social rules. It is also the archetype of a prohibiting rule, i.e. a rule of the general form '(in circumstances C), don't do X', where X stands for a certain type of action or behaviour. The possible qualifier ('in circumstances …') (Hodgson, 2015b) means here that many societies know exceptions to that rule, be it self-defence, euthanasia or war.[4] Despite the obvious benefits of that rule, it is also a rule that needs to be enforced, as evidenced by the ubiquitous presence of the rule in penal law. Of course, whether enforcement is effective –whether punishment is 'sufficiently deterrent' – is a different matter altogether. Suffice to say that the rule is not self-enforcing in the sense that complying with it has as such already positive consequences for the actor.

A closely related type of rule can be termed constraining rule. Such a rule does not exclude certain actions qua action type. Rather, a constraining rule can be said to reduce the *range*, and thereby the number, of admissible and relevant choices from a given set, where range can be conceptualised in terms of the (measurable) characteristics of an action or behaviour compared to (an-)other one(s) in the same choice set (Rosenbaum, 2000). In contrast to a prohibiting rule, the available choice set is therefore not empty; it only becomes smaller compared to a situation without the rule. At the limit, i.e. when the choice set is empty, a constraining rule converges towards a prohibiting rule. Concomitantly, constraining rules also need to be enforced, even though the necessary level of enforcement may prove to be somewhat lower due to the greater range of admissible actions.[5] Which rules belong to this category? Parts of the traffic code comprise constraining rules, a good example being a speed limit. A speed limit does not prevent me from driving at certain speeds; it just limits my choice of speed. From an economic point of view, it appears that property rights have significant constraining elements for both the owner (who cannot do whatever s/he likes with the property) as well as (and

[4] The qualifier 'in circumstances C' may not only comprise exceptions but also the conditions under which a rule is to be applied.

[5] The argument here, defended in more detail in Rosenbaum (2000), is that freedom of choice and therefore constraints limiting freedom of choice can be conceptualised in terms of the range of the characteristics of a choice set. To illustrate the point, a choice set *A* containing five identical white balls offers fewer relevant choices than a choice set *B* containing say, a white ball, a red cube, a green pyramid, a blue tube and a yellow disc. Concomitantly, reducing (constraining) the choice set *A* to only four identical white balls is much less constraining, if at all, than prohibiting the choice of one of the elements in *B*. Note that whether a rule is only constraining or also prohibitive depends not only on substance of the rule but also the choice set to which the rules are applied. A speed limit of 100km/h is not restrictive at all if it applies to cyclists, but would have to be considered prohibitive if it applies to aircraft with much higher minimum speeds.

https://doi.org/10.1017/S174413742100059X Published online by Cambridge University Press

predominantly) for all the non-owners whose freedom of action is limited *vis-a-vis* the owner and his/her property.

The two types of rules discussed so far have in common that their defining element is their being constraints. They are what they are by virtue of what they constrain or prohibit. Moreover, they both come with a qualifier ('in circumstances X') that specifies the conditions under which the rule applies, and, to repeat, constraining and prohibiting rules are the type of rules for which enforcement often appears necessary.

Facilitating and enabling rules

Let me now discuss two (again related) types of rules that also constrain behaviour. However, in contrast to constraining and prohibiting rules, the two types of rules to be analysed now are what they are by virtue of what they seek to achieve, not by virtue of what they seek to prevent, even though they also impose constraints on the behaviour of actors.

Consider the rule that in continental Europe (and many other parts of the world), vehicles are obliged to drive on the right side of the road. Clearly, the essence of this rule is not that it prohibits you from driving on the left side. After all, there are also many parts of the world (and indeed also in Europe), where driving on the left side is equally mandatory. If it was the essence of the rule to make sure that people drive on the right (left) side, then one might ask why this is considered to be better than driving on the left (right) side (and *vice versa*). The point is, of course, that it doesn't seem to matter where we drive provided drivers *agree* on the side.

While also constraining behaviour,[6] rules of the kind just discussed appear to achieve one thing above all others: they facilitate action by coordinating individual behaviour. That is why I shall refer to them in what follows as *facilitating* rules. More generally, facilitating rules can be understood as establishing *conventions*, i.e. agreements on a standard of behaviour, conduct or indeed any other specification. Cases in point with enormous economic significance are the systems of weights and measures, technical norms and standards. One can of course imagine a situation where these problems are negotiated between the actors, for instance each time two vehicles meet on a road, but the costs of doing so in terms of time losses and possible accidents would be enormous. At the same time, the road infrastructure would have to be designed and built in such a way as to be able to accommodate both types of traffic, and this too would be extremely costly.

Couched in these terms, the issue of enforcement looks distinctively different for facilitating rules (Guala, 2015). If confronted with the choice of whether to comply or not, compliance is usually the preferable option – not because non-compliance would be heavily sanctioned, but because compliance itself brings tangible benefits for the actor and non-compliance tangible costs irrespective of any sanctions. A producer of printing paper, for instance, would take considerable risks by offering formats other than A3, A4 or A5. By complying with the industry standards, s/he is on the safe side and does not have to worry about the compatibility of their product with related products.

Against the background of the foregoing discussion, one might still wonder whether rules 'seeking to prevent Y' and rules 'seeking to achieve X' are really so different as to warrant a different terminology. Couldn't they be reformulated into rules 'seeking to achieve non-Y' and rules 'seeking to prevent non-X'? The answer turns, it seems, on what exactly non-X is for a given X. For prohibiting or constraining rules, what we might want to achieve is much less straightforward than what we want to prevent. That is why prohibiting rules are couched in negative terms. Since speeding leads to unsafe driving, it is prohibited rather than calling upon motorists to drive in a safe manner (or prevent accidents, or save lives, etc.). For coordinating rules, the opposite seems to be true. There are often myriad ways of doing things incorrectly (myriad sizes of paper that do not fit into a given printer), but only a few (if more than one) to do things correctly.

[6] If there are just two options, making it mandatory to drive on the right side is tantamount to a prohibition to drive on the left side.

https://doi.org/10.1017/S174413742100059X Published online by Cambridge University Press

In what follows, I will discuss a fourth type of rule, which the above definition of institutions as rules does not fully capture either, namely enabling rules. These rules are located somewhere in between both constraining and facilitating rules without being reducible to either of the latter.

Enabling rules can be characterised by a paradox. Enabling rules constrain more than facilitating rules, but the incentives to comply with enabling rules are even stronger. Why is this so? To begin with, recall what facilitating rules do: they facilitate certain actions and behaviours by solving coordination problems, but these actions and behaviours would still be possible if the rules did not exist. We can imagine a world without traffic rules but with traffic, surely more dangerous and chaotic but nevertheless possible. By contrast, language without rules governing pronunciation, meaning, syntax and grammar would be impossible. That is why we do not understand a foreign language without learning all that, at least in a rudimentary form, and why we have to resort to another type of language (e.g. signs, gestures) if we do not speak the language of the person with whom we want to communicate.

Enabling rules thus differ from facilitating rules insofar as the former do not merely support an activity, these rules constitute this activity in the way chess is constituted by the rules of chess and English is constituted by the rules governing pronunciation, meaning, syntax and grammar of the English language (Searle, 2005). Change the rules of chess and the game is no longer chess but something else. Change the rules of the English language and it becomes something different, first perhaps a dialect and then a different language altogether. This implies, however, that whoever wants to play chess must comply with the rules of chess; otherwise s/he will be ousted from the community of chess players and can no longer play, and whoever wants to communicate in English must comply with the rules of the English language else s/he will not be understood.

These considerations help to explain the above paradox as two sides of the same coin: enabling rules impose far-reaching constraints, but these constraints also ensure their functionality. Nevertheless, even enabling rules are to some extent arbitrary in the sense that very different sets of rules can fulfil the same function(s), as the huge number of languages on Earth demonstrates. Importantly, however, enabling rules rarely come alone; they usually come in sets of various sizes and their components are to some extent interdependent. Again, languages are a case in point in that they comprise vast sets of rules on pronunciation, meaning, syntax and grammar, some of which are indispensable while others are not or only to a certain degree. At the same time, enabling rules are usually, like facilitating rules, coined in terms of what they seek to achieve rather than what they seek to prevent. There is only one way of spelling the word 'correctly' correctly (i.e. according to the rules), but myriad ways of spelling it incorrectly.

The monetary system with its vast set of rules governing not only the issuance of money by the central bank but also the functioning and behaviour of private banks is arguably one of the most important sets of enabling rules in the economic realm. Of course, the monetary system also comprises numerous constraining rules in the above sense, which essentially narrow down the range and nature of choices available to economic actors (and which, therefore, often need to be enforced by a third party). Accounting rules are a case in point. But unlike the rules, which enable the monetary system (those describing the role and functioning of central and private banks), mostly constraining rules are necessary for the well-functioning of the monetary system, but they are not sufficient for its existence. They are thus a kind of auxiliary rule. Unlike other constraining rules, therefore, many of these auxiliary rules are not of intrinsic importance. They would not make much sense outside of the context of, for instance, the monetary system.

The difference between facilitating and enabling types of rules can perhaps best be seen by considering the subset of rules governing the issuance of paper money, i.e. those rules which assign the function of counting as money to certain physical objects. Accordingly, only pieces of paper with a specific design printed on them and produced and handed out under the authorisation of the central bank count as money. Private banks and other economic actors are permitted to use banknotes as a means of payment and for other purposes, but they are not allowed to reproduce or issue them, the latter being an auxiliary prohibiting rule. This auxiliary rule is important for a well-functioning monetary system, but has no meaning outside of the latter.

https://doi.org/10.1017/S174413742100059X Published online by Cambridge University Press

The preceding example illustrates another feature of enabling rules. Enabling rules of the kind described in this section allow individuals or groups to create institutional facts (Searle, 1995, 2005). Institutional facts are social facts, i.e. facts involving the collective intentionality of two or more agents, which can only exist given certain institutions and the commitment of agents to these institutions. For example, specific entries on the balance sheet of a bank (demand deposits) count as money and can therefore be used to make payments, but only insofar as everybody concerned believes this to be the case and behaves accordingly, including the application of the relevant rules. Institutional facts of this kind are so pervasive in human societies that we hardly recognise them as such.

For the purpose of the present discussion, it is important to keep one distinction in mind though. When we speak of money, for instance, we can refer either to the monetary systems (i.e. the rules governing the creation and issuance of money), or we may refer to a concrete instantiation of this system in the form of a 10€ bill in my wallet. The latter is not an institution but an institutional fact in the above sense. A piece of paper with certain characteristics as determined by the monetary system becomes an instantiation of money only if this fact is socially recognised.

Role, actor and status

The typology of rules that has been developed above is incomplete because unlike in the case of the three other rules discussed, it doesn't seem to be fully clear what form enabling rules take. Secondly, the relationship between rules and actors needs to be further clarified. Are actors just automatons of some kind whose behaviour in the context of an institution is fully determined by the institution's rules or is there more? As I shall argue, both gaps are related and need to be filled in order to understand both institutional development and inertia. In this section, I will therefore examine what Searle (1995, 2005) has called status functions. Searle's account focuses predominantly on the roles assigned to actors, but as shown by Faulkner and Runde (2009, 2013), similar concepts can also be fruitfully applied to immaterial (technological) objects.

At the outset, it should be pointed out that the typology of rules developed above distinguishes at best implicitly between (different types of) actors, or more precisely, between the status and the roles assigned to different actors where status is provisionally to be understood as the position of a person in relations to others (more on this below) and the notion of role relates to the specific functions of a person. This situation is unsatisfactory insofar as the four types of rules (institutions) generally do not address each possible actor, but only certain actors or actors with a certain status or role. So the rules governing the creation of money address (central) bankers or customers of banks, but not infants, and traffic rules concern only drivers, pedestrians and the like but not somebody staying at home. Thus, it is usually by virtue of having a specific role or function that a rule applies to an actor. Once the person in question has no longer the role or status of employee or manager, the rule doesn't apply anymore.

Furthermore, most persons have multiple roles and statuses. They are an employee and car drivers, or they are the president of a sports club and a pensioner. Consequently, multiple sets of institutions apply to them, not necessarily all of them all the time but at least with a certain regularity. But how do actors get a role or a status and how can it be that such a status confers to the actor power and influence? In the account provided by Searle (1995, 2005), three elements are important in order to understand what is going on: collective intentionality, the assignment of functions, which I take to be broadly equivalent to roles as defined above, and status functions.

Collective intentionality 'covers not only collective intentions but also such other forms of intentionality as collective beliefs and collective desires. One can have a belief that one shares with other people and one can have desires that are shared by a collectivity' (Searle, 2005). So understood, collective intentionality is the basis of human cooperative behaviour, of doing things together rather than just in parallel or at the same time.

By the assignment of functions, it is to be understood that '[h]uman beings have a capacity … to impose functions on objects where the object does not have the function, so to speak, intrinsically but

https://doi.org/10.1017/S174413742100059X Published online by Cambridge University Press

only in virtue of the assignment of function' (Searle, 2005). Tools are a case in point, where an object is assigned a function. Perhaps Searle goes too far when he claims that the assignment of a function can always be undertaken irrespective of the properties of the object. So let us settle on the interpretation that at least most artificial objects that surround us do not perform their functions randomly, but because they were designed with a specific purpose in mind. On the other hand, human beings can also assign functions to objects or states of play in ways that are only loosely related to their physical properties, if any. By looking at these objects or states of play, it is therefore far from evident what their purpose and function might be. A text on a piece of paper may be a poem or a contract. A group of people coming together may be a religious congregation or a parliament.

By the assignment of status, Searle (2005) finally means 'a special kind of assignment of function where the object or person to whom the function is assigned cannot perform the function just in virtue of its physical structure, but rather can perform the function only in virtue of the fact that there is a collective assignment of a certain status, and the object or person performs its function only in virtue of collective acceptance by the community that the object or person has the requisite status'.

To summarise, Searle posits that institutional facts (facts which can only exist given certain institutions) 'typically require structures in the form of constitutive rules X counts as Y in C and that institutional facts only exist in virtue of collective acceptance of something having a certain status, where that status carries functions that cannot be performed without the collective acceptance of the status'. Concomitantly, his account explains how actors obtain a status that makes certain rules applicable to some actors and not to others. A set of rules applies only to an actor if it has been collectively accepted that the actor has a status that makes that set of rules applicable to the actor.[7]

For Searle (2005), an institution, then, is any system of constitutive rules of the form 'X counts as Y in C'. A comparison with the four types of institutions above suggests, however, that the form of the first three is clearly different. Neither a prohibiting rule nor a constraining rule nor a facilitating rule resembles Searle's constitutive rule. I therefore suggest treating these rules as institutions, which are different in kind from Searle's. By contrast, enabling rules and constitutive rules appear to be similar if not synonymous. I shall argue in a moment though that an exclusive account of enabling rules in terms of constitutive rules is insufficient.

Of course, many enabling rules do take the form suggested by Searle. A certain configuration of the pieces on the chessboard counts as checkmate. However, I surmise that these rules are not the only rules constituting chess or the monetary system for that matter. There are also rules describing the admissible and non-admissible moves on the chessboard. Do such rules also have the format 'X counts as Y in C'? Admittedly, it seems possible to reformulate constraining or prohibiting rules in such a way that they resemble Searle's format ('moving a pawn one field forward counts as an admissible move in chess'). However, the added benefit seems questionable.

In criticising Searle's account, Hindriks and Guala (2015a, 2015b) have argued that Searle's (1995, 2005, 2015) constitutive rules can be reduced to regulative rules and that therefore there is nothing special about Searle's account. Notions such as property only serve as a shortcut for a set of rights (and obligations). The former authors appear to be right in the sense that regulative rules are necessary to describe (or define) notions such as property (or POTUS) but they are clearly not sufficient. Notions such as property go hand in hand with being an owner (or a neighbour) and notions such as voting go hand in hand with being a congresswoman or a voter.

Importantly, these are functions which cannot be reduced to a set of rights and obligations (such as the function of a guard who closes and opens the barriers at a level crossing), they also include a broadly shared understanding of what it means to play the role of president (or owner) well (rather than just going through the motions as it were). Although the objectives of being President are surely more multidimensional than the objectives of a chess player, who wants to win the game by

[7] To be sure, not all social facts may require enabling rules of the kind discussed here, but many important social facts do. Buying a house is one of the latter; offering my partner some blueberries I just picked as a gift is perhaps one of the former, even though one might argue that even giving a gift requires some basic notions of property.

https://doi.org/10.1017/S174413742100059X Published online by Cambridge University Press

checkmating the opponent, there can be no doubt that ‘playing the Presidential game’ correctly is one thing, playing it successfully is something quite different and goes way beyond knowing and obeying the rules of the game, and not only that: there are usually also rules which indicate how to play the presidential or any other game well while whole books have been written about how to play chess well. After all, it would be a rather pointless exercise to move around the chess pieces in accordance with the rules but without knowing how and with what purpose or goal this has to be done. This is arguably an important element of what in Searle (2005) is referred to as social intentionality and where the latter connects with the above debate on the understanding of institutions.

These considerations suggest that in particular enabling rules must comprise (or go hand in hand with) some notions as to why and with what purpose a set of rules exists and is used. Actors must understand why they are doing what they are doing, including the specific kind of role assigned to them in relation to the roles assigned to others, and how they can increase their chances of being successful, i.e. achieve what they intend to achieve. It is not sufficient to know the rules of the game as it were, but the players need to know how the game works, which moves may or may not be advisable given a certain constellation of the pieces on the chessboard, which opening moves to make and how these relate to the overall strategy a player is pursuing.

This, I surmise, is part and parcel of having a certain status and playing a specific role and it is also in a way inseparable. A certain status and some commonly shared expectations about what it means to have that status and to play that role well go hand in hand. Actors are assigned specific roles precisely because we (the voters, the shareholders, etc.) expect them to play that role well, not only or even primarily because they know the rules. Note that playing a role is different from conscious goal seeking in that the former includes a predisposition to act in one way rather than another, not a conscious decision. A role is in this sense not only determined by the current status but also by history, i.e. current and past positions (Cardinale, 2018). Having a role is not say that conscious decisions are replaced by those prescribed by the role. The point is rather that, within the context of that role, certain sets of actions are precluded (or endorsed) *a priori* and the conscious choice takes place within these sets. This feature of institutional roles has also implications for the understanding of mental models, to which I shall turn now.

3. Institutions and mental models

The concept of mental model

It is widely acknowledged that institutions must be represented mentally in order to be effectively impacting behaviour. Actors can only follow a rule of which they are aware, and they can only decide not to follow a rule if they are aware of the rule either (Lawson, 1994; Searle, 2005). At some point, of course, conscious rule following may turn into a habit, in particular when agents repeat actions frequently (Dequech, 2013), but acknowledging this does not invalidate the general principle and a habit will usually be rendered conscious and subsequently abandoned if it becomes counterproductive as it were.

This section argues that in particular for enabling rules, the necessary notions on the part of the actors as to why and with what purpose a set of rules exists and is used can best be understood by means of the concept of mental model. In what follows, ‘[m]ental models are conceived of as a cognitive structure that forms the basis of reasoning, decision making, and, with the limitations also observed in the attitudes literature, behavior’ (Jones *et al*., 2011). Denzau and North (1994) have introduced this concept, which dates back to the work of the Scottish psychologist Craik (1967), into institutional analysis.

Generally speaking, mental models are subjective representations of objective facts and subjective representations of subjective (or social) facts (institutions, power structures, etc.), the latter being the product of social interaction, communication and agreement as outlined in the previous sections. These representations are structured in the sense that different elements are ordered by relations such as assumed causalities or dependencies. Such causalities can be physical (if I accelerate my car too

https://doi.org/10.1017/S174413742100059X Published online by Cambridge University Press

much, it will skid on a wet road) or social (harsher punishments lead to a decrease in crime). Thus, the facts represented by mental models do not only include accounts of objects and events but also, as I would like to emphasise, of presumed causal relationships between objects and events, including the assumed effects of our own actions and those of others in a specific context.

Mental models are similar to architects' models or to physicists' diagrams in that their structure is analogous to the structure of the situation that they represent. That is why mental models are a subclass of cognitive representations but not every cognitive representation is also a mental model. Neither are my memories of the face of a person I may have seen on the street in this sense a mental model (there is no structure here), nor a formal theory such as the IS/LM model (its structure is not analogous to the situation it represents), although the memory of a person's face may be part of a mental model that indicates how to interact with that person.

Despite being subjective, mental models are not idiosyncratic but ripe with shared meanings and interpretations (Lewis, 2005). People face always and anywhere social structures. '[I]n attempting to divine the significance of price signals, for example, people are able to transcend a purely subjective and therefore potentially arbitrary and idiosyncratic viewpoint only by drawing on the traditional conceptual schemes they share with other members of their society' (Lewis, 2005). Some mental models may be rather simple and straightforward, others complex and intricate, depending on the role of the person concerned. So the mental model of a rail passenger of the rail system, while still important to find the way from A to B, is bound to differ substantially from that of a train driver. For the passenger, such a model would focus on possible endpoints of journeys and the need to change trains in between. For the train driver, it is more important to know the technical peculiarities of the line between stations (speed limits, gradients, the position of signals, etc.) and how to react to these peculiarities taking into account the technical specificities of the train.

In short, mental models embody the knowledge about how the world around us, with which we interact, functions and responds in turn to our actions. Therefore, mental models guide decisions and choices by identifying the options which, as we believe, help us achieve our goals while excluding those that run counter to our objectives. Thus they help to form expectations about the environment (Denzau and North, 1994; Holland *et al.*, 1986) by embodying knowledge and beliefs about causal relationships, but they are not synonymous with expectations, except in a very generic sense. In parallel, mental models also filter information (Loasby, 2001).

This understanding of mental models contrasts with the view advanced by Cubitt and Sugden (2003) and Sugden (2015), who have argued that institutions as rules do not have to be represented by (or within) mental models, but as symbols. However, while symbols often serve as shortcuts for a verbal description of the rule, they do not convey any information beyond that and so cannot substitute a mental model as understood here. On the contrary, while symbols are means to condense the informational content to a bare minimum, mental models provide information that goes way beyond the rule itself. Thus, a stop-sign at a crossroad signifies that any vehicle approaching that crossroad needs to stop in front of the stop-sign. It does not say with which speed to approach the stop-sign though (doing it too fast may irritate other drivers), or how long to stop or what to do while stopping. All this, I surmise, is part and parcel of what a driver has in mind when approaching a crossroad with a stop-sign.

Several additional features of mental models should be noted. First of all, while mental models are in our cognitive system, they are not hard-wired in our brains but are culturally transmitted via processes of learning through imitation and formal and informal schooling. This implies that mental models may undergo changes in the course of time, be it because we learn new or different things, be it because the feedback we get from our interaction with the environment prompts modifications to our mental models. Negative feedback in particular is likely to lead to modifications as it suggests that something is wrong with a mental model while positive feedback confirms the model. However, modifications are neither automatic nor predetermined but thinking and reasoning serve as internal manipulators (Johnson-Laird, 2004). This also suggests that mental models operate at different levels of consciousness and awareness. For routine tasks, they operate by and large in the background and are

https://doi.org/10.1017/S174413742100059X Published online by Cambridge University Press

taken for granted, while becoming the subject matter of full discursive awareness when things go wrong or when the actor is confronted with a hitherto unknown problem or situation.

Secondly, mental models embody explicit knowledge (as opposed to tacit knowledge (Polanyi, 1967)).[8] In other words, the knowledge that they comprise can be spelt out and communicated (and hence taught and learned). Tacit knowledge is often complementary to the explicit knowledge contained in mental models (Nonaka, 1994). Tacit knowledge must also be learned, but it is acquired through (repeated) practice of the activity in question, rather than being taught. Once tacit knowledge has been successfully acquired, however, no conscious effort is needed to activate it. Our body automatically knows what to do so to speak. While a cyclist can certainly describe in some detail how to ride a bicycle, no account would suffice to enable a non-cyclist to take a bike and start riding it.[9] But once a person has learned to ride a bike, s/he will never forget it.

Thirdly, any mental model is necessarily partial in the sense that it does not provide a comprehensive description of reality irrespective of its purpose and context (and does not intend to) as already the above example of the different mental models of rail passengers and train drivers suggests. However, this is not achieved via abstraction: 'We cannot start with a complex reality, and choose how to simplify it by removing some connections: that is a cognitive impossibility. Instead, knowledge has to be constructed by building up connections'(Loasby, 2001).

Fourthly, mental models are of particular importance for enabling rules, which help us constructing or doing things. Why is this so? As indicated above, it is not sufficient, for instance, to know the rules of chess in order to play the game well. The rules of the game as laid down in its constituting rules do not tell players what to do; they inform the player only about the admissible moves and the configurations of pieces that have specific significance, such as checkmate. But the player also needs to have an idea of how the game functions, what the objectives of each player are (or should be) depending on the state of play, which strategy to use and which tactics to apply. Hence, mental models provide the extra information and knowledge that is necessary in order to not only comply with, but to purposefully apply, institutions. Without predetermining behaviour (and thus making it devoid of agency), the roles attributed to, and acknowledged by, agents in a given institutional setting and as reflected in their mental models make some actions more likely than others. This is what Cardinale (2018) has termed the orienting function of institutions: they provide structure but simultaneously also induce actors to follow some possible actions rather than others (Cardinale, 2018). In some situations, groups of people share specific mental models (Cooke *et al.*, 2000) which may differ only insofar as different team players have different but related roles. Of course, there is no guarantee that mental models always match as they were. In particular, where the underlying institutions are only loosely specified and formalised, mental models may not fully match and these differences may occasionally lead to uncertainty (Wrenn, 2006) and conflicts.

Crucially, and following from the foregoing remarks, many institutions, despite being formalised, cannot be seen as being independent of their mental representations as Denzau and North (1994) seem to insinuate occasionally. Therefore, even if the wording of the rules was the same as it is the case with formalised rules, the knowledge of how to use them is embedded in the associated mental models and changes with the latter. A set of rules can lead to a different outcome if our understanding of how to apply and operate with these rules and our associated interests and goals change. This has an important ramification. Society cannot, in a deep sense, be wrong about its institutions, only some individuals can to the extent that their mental models differ from those of the other members of society, prompting them to 'misbehave'. Concomitantly, when the mental models of an increasing number of people undergo changes that go into the same direction, then, inevitably, the institutions

[8] In Wrenn's (2006) exposition, the main building blocks of mental models are instincts, habits and patterns of behaviour. This seems to downplay somewhat the importance of knowledge.

[9] In contrast to Faulkner and Runde (2013), I would argue though that the rules of grammar are not tacit knowledge in this sense because it is possible, at least in principle, to ask a competent speaker of English to enunciate for instance the correct conjugation of the verb 'to be' in a way which allows a person learning English to reproduce it. Tacit knowledge, by contrast, is often brought to light using metaphors because it is not possible to describe it (Nonaka, 1994).

https://doi.org/10.1017/S174413742100059X Published online by Cambridge University Press

complemented by these mental models will also change because the way actors behave and thus reproduce these institutions will change. The development of languages in the course of time illustrates this point. While medieval and modern English are surely different in terms of grammar, syntax, etc., neither can be said to be correct or incorrect English. Generally speaking, therefore, when institutions (and hence the social structure to which they give rise) are reproduced through their everyday use and application, this is not merely an exercise in duplication of identical things, but a reconstruction shaped and possibly modified by the experience of actors.

It is for precisely this reason that the notion of reproduction in realist philosophy does not appear to be uniform, but addresses different aspects of the same problem. It can mean (i) to act in accordance with a (set of) rule(s), (ii) to do so on the basis of a mental model which incorporates and complements the rule(s) and/or (iii) to get positive feedback with respect to the action in general and the rule and the mental model in particular. Without being able to elaborate on the issue, it seems that all three aspects are important for the reproduction of institutions, as none of them achieves that goal without the others. But taken together, they also imply that institutions are not static over time. They change and evolve, perhaps imperceptibly so, at any point in time, but nevertheless visible across longer periods.

Before concluding this section, I shall address a possible ontologically grounded critique of the above argument. Fleetwood (2019) distinguishes between formal rules and norms (or informal rules): following the former requires always a conscious decision, following the latter does not or not always as some informal rules are also followed unconsciously. Importantly, however, being formal means that a rule is *located in an artefact* such as a book or a sign (e.g. a sign saying that hard hats have to be worn on a building site). Such rules, Fleetwood (2019) argues, exist therefore separately from agents although they may be memorised and later remembered by agents up to a point when following them becomes an unconscious process, quasi-independent of the formalisation. Informal rules, by contrast, are said to be ‘located in agents’ cognitive systems as memories of past actions’ (Fleetwood, 2019: 26). Accordingly, they cannot exist separately from agents themselves. Thus one cannot argue, as one referee put it, that institutions are formal rules = social stuff = ‘out there’, and simultaneously argue that institutions are formal rules = cognitive stuff = ‘in here’, i.e. somehow represented by, or even located in, mental models. According to this view, formal rules can at most be remembered.

There are three interrelated reasons – one perhaps somewhat weaker, the others stronger – why it is not so easy to say where formal institutions actually reside (although it is not wrong to say that they also reside in artefacts). The first argument runs as follows. Suppose that all artefacts in which a formal rule (or set of formal rules) is embodied disappear from one day to the next, say all legal texts, text books, etc. Fleetwood’s argument taken literally suggests that this would also apply to the institution itself. After all, if a rule is embodied in a book, where does the rule go, so to speak, when the book disappears? But is that really the case? After all, there would still be all those agents (in the case of the legal system, the lawyers, judges, legal scholars, etc.) who *collectively* know much of the rules and principles that make up the legal system. If they do, then all these actors could begin a collective exercise aimed at reformalising the legal system, i.e. at (re)building and recreating all the artefacts that have been destroyed or have disappeared. This may take some time and not cover each detail, but it does not seem impossible. If the example is correct, it therefore suggests that formal rules are not exclusively contained in artefacts. One could even argue that the purpose of formalisation is precisely to ensure reliability and consistency, for instance in teaching, and to make rules more independent of the vagaries of human memorisation and verbal communication, in particular in contexts such as law or politics where precision and unambiguity may be a matter of life or death. Hence, formalisation is arguably (only) a secondary and auxiliary phenomenon.

This brings me to my second argument. Before it becomes possible to formalise a rule, the rule must first be thought through and possibly even developed in individual and collective reflexive processes. While influenced by language, logic, etc., these processes are nonetheless located in the minds of the agent(s) involved. This is similar, or so it seems, to writing down the ideas expressed in this

https://doi.org/10.1017/S174413742100059X Published online by Cambridge University Press

article. The author must first develop them in his or her mind before they can then be articulated – in this case – in the English language. While the process of writing them down supervenes on the grammar, syntax, and semantics of the English language and the author's knowledge and mastery of that language, these ideas are certainly not determined by the language with which they are expressed. In other words, formalisation presupposes clarification and conceptualisation, which are in turn cognitive processes. A telling example in this regard is the introduction of the meter as the base unit of length in the International System of Units (SI). This was the result of long political and scientific debate, first in France, then in other countries, which only when –temporarily concluded – lead to the creation of an artefact: the prototype meter bar. Even a formal rule exists first in our minds before it will undergo formalisation and then exists also outside our minds.[10]

The third argument takes up a point made above. There, it has been argued that rules/institutions continue to exist even in the absence of their concrete and uninterrupted *realisation*. However, when they are realised, then this realisation (take, for instance, the text you are currently reading) also embodies the rules, albeit in an implicit and incomplete form. While this implicit form may not suffice to reconstruct all relevant rules (as a short English text may not suffice to reconstruct English grammar, syntax, etc., in its entirety), it surely does to some degree. I take it that this form of embodiment in practice is therefore ontologically different from an artefact such as a dictionary or a grammar book, which seeks to represent rules in an explicit and comprehensive form.

The distinction between formal and informal rules and their respective locations is nevertheless an important one, pointing as it does to the idea that formal institutions have also a significant cognitive role in the sense that, for instance, the legal systems with its vast panoply of codified substantive and procedural rules shape and direct our thinking about legal issues and thus is part and parcel of cognition (Gallagher, 2013; Gallagher and Crisafi, 2009). This suggests not only that there is always a reciprocal relationship between institutions and how agents make sense of them and use them, but also that cognition cannot be comprehensively conceptualised as something that takes place exclusively in agents' minds. It encompasses their interaction with other agents and the inanimate environment (de Bruin and Kästner, 2012), including all the clever artefacts from notebooks to quantum computers designed to support our memory and our computational abilities.

This being said, it is time to wrap up. A useful starting point for a representation of what is going on can be found in Aoki (2015) who sees the creation and reproduction of institutions as an interactive and self-referencing process where behavioural choices create structures (institutional facts) which either directly or indirectly (via public representation) shape behavioural beliefs[11] and mental models, which, in turn, motivate behaviour. Using Aoki's graph for a representation of the key relationships between the various dimensions does not imply that his game-theoretic approach is hereby endorsed and it is for precisely this reason that several aspects of the graph have been modified. In contrast to Aoki (2015), I would argue that people's interaction with institutions and with other agents may also impact the behavioural beliefs and mental models they hold and that these beliefs and models then also shape the public representation and formalisation of institutions. Thus there is hardly if ever a one-way relationship between behavioural choices, the formalisation and representation of institutions and mental models.

From this, it follows that behavioural beliefs and mental models do not only motivate and inform behavioural choices, but the latter also inform our beliefs, be it directly (feedback) or indirectly (emulation), and this is also the case for the structures (i.e. the relations between people with their various roles, rights and obligations) which emerge directly and indirectly from behavioural choices. Concomitantly, the public representation and formalisation of rules is not only likely to induce beliefs and mental models. Our beliefs and mental models also shape the public representation and

[10] Note that this argument primarily concerns the emergence of a formal rule. Once such a rule is established, it resides both in our minds and in various artefacts, but not equally so for everyone. Those who encounter the rule for the first time may do so with its embodied form, i.e. a form located in an artefact such as a lawbook or a measuring tape, but they also might be confronted with realisations of the rule.

[11] According to Aoki (2015), behavioural beliefs are expectations of an agent regarding what the other agents are doing and how they will react to what s/he does.

https://doi.org/10.1017/S174413742100059X Published online by Cambridge University Press

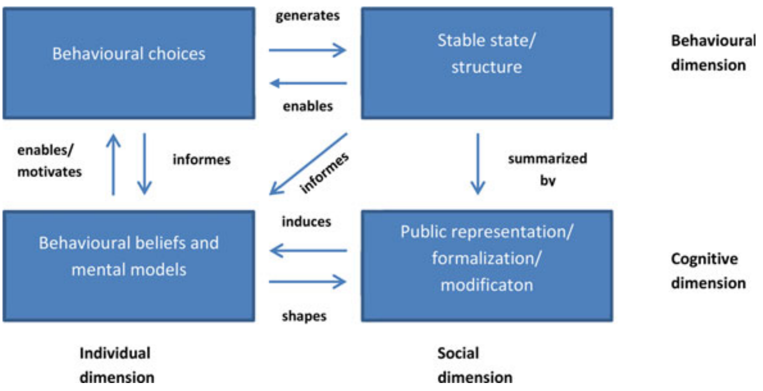

Figure 2. (Based on Aoki (2015)).

interpretation of institutions. At the end of the day, we see therefore a much more complex pattern of interaction between mental models, behavioural choices, structures and the representation and formalisation of institutions in which these various elements are intimately intertwined. Concurrently, both institutional inertia and change are to an important degree endogenous phenomena. While some of the relationships depicted in Figure 2 are likely to stabilise the institutional setting via reinforcing feedback mechanisms, others may induce instability and hence put the institutional setting on a path of dynamic change.

4. Subject matter of mental models and policy implications

What is the content of mental models of economic agents and isn't the notion of mental model similar to the concept of rational expectations, which are usually modelled as agents knowing how the economy operates and responding based on this knowledge? The answer to the latter question, I presume, is Yes and No. It is Yes because there are certain similarities between rational expectations and mental models: Both involve an understanding of how the economy works. It is No because there are nevertheless important epistemological and ontological differences.

In epistemological terms, the difference is that we cannot know how the real economy looks like, only a very much simplified understanding is possible. But as pointed out above, such an understanding is not achieved purely via abstraction but inductively. I would even go further than that and claim that much of neoclassical theorising is not based on abstractions but on idealisations. After all, it is not the case that *homo oeconomicus* all of a sudden appears after we have stripped real people of their individual features and thus have abstracted from their distinguishing properties. Concomitantly, Austrian economists would emphasise that our beliefs, even where accurate in some respects, are usually partial and often fragmentary (e.g. most proponents of AE would be reluctant to assume that someone who has knowledge of some body of true propositions will automatically know the deductive consequences of those propositions) (Runde, 2002). This, too, distinguishes the rational expectations approach from heterodox points of view.

In ontological terms, the difference is that, as the economy consists to an important extent of socially reproduced structures including Searle's institutional facts, what they are and what we believe them to be is not independent but closely intertwined. While rational expectation theorists would maintain that there is a correct model of the economy about which we may hold correct or incorrect beliefs, the transformational model of social institutions implies that social structures and the beliefs we hold about them are not disjunctive, or, to put it more bluntly, if people believe that the economy functions, say according to the neoclassical model, then it will leave an imprint as beliefs shape expectations and thereby behaviour. If people believe instead that it functions according to a Keynesian model, then this belief too will leave an imprint on how the economy functions.

https://doi.org/10.1017/S174413742100059X Published online by Cambridge University Press

Claims of this sort are less far-fetched than they may seem at first sight. As MacKenzie (2008) has shown in detail, neoclassical theories of finance were not external analyses of pre-existing economic phenomena, but rather normative stances which became parts of economic processes and thus altered financial markets fundamentally. By propagating a model of financial markets, which guided – once internalised by economic actors into their mental models – their behaviour in ways consistent with the theory, neoclassical theories of finance became quasi self-fulfilling prophecies and their (empirically underpinned) 'truth' did not reflect some deeper knowledge about the functioning of financial markets but rather the transformation of these markets in the light of the normative views developed by theorists and subsequently held by market participants.

It is of course difficult to make generalisations about the content of people's mental models and in particular their economic mental models. The research that comes closest to such a type of investigation are studies which seek to examine the economic knowledge of consumers or the population at large (e.g. Wobker *et al.*, 2014 or Bucher-Koenen and Lusardi, 2011; Jappelli, 2010). However, these studies take as their starting point the view that there is a settled body of economic knowledge, which is ontologically independent of the economy and its structures and institutions. As argued above, this view is problematic.[12] So rather than asking about imputed causalities, the investigation of mental models requires, or so it seems, an approach that is open with respect to the possible causal relationships and structures.

Surveys which have been undertaken in the context of the transformation of former centrally planned economies in Central and Eastern Europe appear to be less biased in that respect. Importantly, analyses based on such surveys suggest that differences between mental models about the functioning of market economies explain to a significant extent differences of reform success (Rosenbaum, 2001), thereby underpinning the above claim that institutions and their representation are interdependent.

Tangentially to these findings, there is evidence to suggest that broad trends in economic policy making are to a lesser extent than perhaps commonly though (or hoped) shaped by debates in academic circles (and therefore universities), and to a greater extent by paradigms and public sentiments (Campbell, 1998) containing beliefs and convictions held by policy makers and the general public, respectively, about how the economy works. These are (at least) as much influenced by deeply ingrained socio-cultural beliefs and political convictions as they are by scientific findings and evidence.

5. Conclusions

This essay has argued that institutions as enabling socially accepted rules are not only self-reinforcing, they are also often complemented by complex culturally transmitted mental models. Both features are bound to prevent rapid changes and thus manifest themselves in what I have termed institutional inertia. The former implies that agents get positive feedback from applying such rules, while the latter implies that change presupposes or necessitates learning processes, which are likely to take time and therefore slow down modifications to existing institutions as well as the adoption of new rules.

Institutional inertia, so conceived, cannot be construed as entirely negative though – quite the opposite. Enabling institutions form the backbone of the structures that make up our societies, and their unique role is not least due to the fact that they do not need an enforcer and therefore do not raise the question of who enforces enforcement. Since these institutions must also be firmly anchored in peoples' minds and mental models, any rapid change is bound to undermine their ability to enable and constrain actions. Thus, their stability is a precondition for their social acceptance and transformative reconstruction.

If mental models are the bearers of many institutions in the way argued above, i.e. in the sense of being essential complements, then institutional change presupposes a change of the corresponding mental models, not only of the corresponding (formal) rules. Such a change is likely to be gradual,

[12] The questions asked in these studies are rather heterogeneous and comprise technical concepts as well as empirical facts and theoretical notions. It is of course primarily the latter that raise ontological issues.

https://doi.org/10.1017/S174413742100059X Published online by Cambridge University Press

at least in the sense that not all holders of a specific mental model will simultaneously switch to another mental model, or at least a modification thereof. Seen from this perspective, institutions are not only stable because they are useful (functional argument) but because they are represented in mental models, which have been learned and would need to be replaced by other mental models or at least significantly modified if an institution is going to be changed. Irrespective of whether there are more efficient alternatives or not, self-enforcing institutions, in particular, are therefore inherently stable. Or, to put it differently, they will resist change and therefore exhibit inertia.

Moreover, it is not just the institutions that tend to resist change, also the vast amount of institutional facts, which has been created on the basis of these institutions, arguably achieves after some time a quality which brings them close to physical facts (and not only because of the physical traces these institutional facts leave behind all over the planet). These facts themselves exert a normative force that consists in having shaped social interaction over a sufficiently long period.

References

Aoki, M. (2015), 'Why is the Equilibrium Notion Essential for a Unified Institutional Theory? A Friendly Remark on the Article by Hindriks and Guala', *Journal of Institutional Economics*, **11**(3): 485–488, doi: 10.1017/S1744137415000090

Bucher-Koenen, T. and A. Lusardi (2011), 'Financial Literacy and Retirement Planning in Germany', *Journal of Pension Economics & Finance*, **10**(4): 565–584.

Campbell, J. L. (1998), 'Institutional Analysis and the Role of Ideas in Political economy', *Theory and Society*, **27**: 377–409, doi: 10.1023/A:1006871114987.

Cardinale, I. (2018), 'Beyond Constraining and Enabling: Toward New Microfoundations for Institutional Theory', *Academy of Management Review*, **43**(1): 132–155, doi: 10.5465/amr.2015.0020

Cooke, N. J., E. Salas, J. A. Cannon-Bowers and R. J. Stout (2000), 'Measuring Team Knowledge', *Human Factors and Ergonomics Society*, **42**(1): 151–173, doi: 10.1518/001872000779656561

Craik, K. J. W. (1967), *The Nature of Explanation*, Cambridge [etc.]: Cambridge University Press.

Cubitt, R. P. and R. Sugden (2003), 'Common Knowledge, Salience and Convention: A Reconstruction of David Lewis' Game Theory', *Economics and Philosophy*, **19**(2): 175–210. 10.1017/S0266267103001123.

de Bruin, L. C. and L. Kästner (2012), 'Dynamic Embodied Cognition', *Phenomenology and the Cognitive Sciences*, **11**(4): 541–563, doi: 10.1007/s11097-011-9223-1

Denzau, A. T. and D. C. North (1994), 'Shared Mental Models: Ideologies and Institutions', *Kyklos*, **47**(1): 3–31.

Dequech, D. (2013), 'Economic Institutions: Explanations for Conformity and Room for Deviation', *Journal of Institutional Economics*, **9**(1): 81–108, doi: 10.1017/S1744137412000197

Faulkner, P. and J. Runde (2009), 'On the Identity of Technological Objects and User Innovations in Function', *Academy of Management Review Academy of Management*, **34**(3): 442–462, doi: 10.5465/AMR.2009.40632318

Faulkner, P. and J. Runde (2013), 'Technological Objects, Social Positions, and the Transformational Model of Social Activity', *MIS Quarterly*, **37**(3), pp. 803–818. Available at: http://aisel.aisnet.org/misq/vol37/iss3/9%5Cnhttp://misq.org/misq/downloads/download/article/1034/%5Cnhttp://misq.org/misq/downloads/%5Cnhttp://aisel.aisnet.org/misq/vol37/iss3/.

Fleetwood, S. (2008), 'Institutions and Social Structures', *Journal for the Theory of Social Behaviour*, **38**(3): 241–265, doi: 10.1111/j.1468-5914.2008.00370.x

Fleetwood, S. (2019), 'Re-visiting Rules and Norms', *Review of Social Economy*, doi: 10.1080/00346764.2019.1623909.

Gallagher, S. (2013), 'The Socially Extended Mind', *Cognitive Systems Research*, **25–26**: 4–12, doi: 10.1016/j.cogsys.2013.03.008

Gallagher, S. and A. Crisafi (2009), 'Mental Institutions', *Topoi*, **28**(1): 45–51, doi: 10.1007/s11245-008-9045-0

Guala, F. (2015), 'The Normativity of Institutions', *Phenomenology and Mind*, **9**: 118–128. doi: 10.13128/phe_mi-18157.

Hindriks, F. and F. Guala (2015a), 'Institutions, Rules, and Equilibria: a Unified Theory', *Journal of Institutional Economics*, **11**(03): 459–480, doi: 10.1017/S1744137414000496

Hindriks, F. and F. Guala (2015b), 'Understanding Institutions: Replies to Aoki, Binmore, Hodgson, Searle, Smith, and Sugden', *Journal of Institutional Economics*, **11**(03): 515–522, doi: 10.1017/S1744137415000120

Hodgson, G. M. (1988), *Economics and Institutions: A Manifesto for a Modern Institutional Economics*. Cambridge: Polity Press. Available at: http://www.worldcat.org/title/economics-and-institutions-a-manifesto-for-a-modern-institutional-economics/oclc/818199477&referer=brief_results (Accessed: 3 October 2017).

Hodgson, G. M. (2006), 'What Are Institutions?', *Journal of Economic Issues*, **40**(1): 1–25, doi: 10.1080/00213624.2006.11506879

Hodgson, G. M. (2015a), 'Much of the "Economics of Property Rights" Devalues Property and Legal Rights', *Journal of Institutional Economics*, **11**(4): 683–709, doi: 10.1017/S1744137414000630.

https://doi.org/10.1017/S174413742100059X Published online by Cambridge University Press

Hodgson, G. M. (2015b), 'On Defining Institutions: Rules Versus Equilibria', *Journal of Institutional Economics*, **11**(3): 497–505, doi: 10.1017/S1744137415000028

Holland, J. H., K. J. Holyoak, R. E. Nisbett and P. R. Thagard (1986), *Induction: Processes of Inference, Learning, and Discovery*. 5th print. Cambridge Mass. [u.a.]: MIT Press. Available at: https://www.worldcat.org/title/induction-processes-of-inference-learning-and-discovery/oclc/246419184&referer=brief_results (Accessed: 24 October 2017).

Jappelli, T. (2010), 'Economic Literacy: An International Comparison', *The Economic Journal*, **120**(548): F429 –F451.

Johnson-Laird, P. N. (2004), 'The History of Mental Models', in K. I. Manktelow, M. C. Chung (eds), *Psychology of Reasoning: Theoretical and Historical Perspectives*, London: Psychology Press, pp. 179–212.

Jones, N. A., H. Ross, T. Lynam, P. Perez and A. Leitch (2011), 'Mental Models: An Interdisciplinary Synthesis of Theory and Methods', *Ecology and Society*, **16**(1), doi: 10.5751/ES-03802-160146

Lawson, C. (1994), 'The Transformational Model of Social Activity and Economic Analysis: A Reinterpretation of the Work of J.R. Commons', *Review of Political Economy Edward Arnold*, **6**(2): 186–204, doi: 10.1080/09538259400000009

Lawson, T. (2003), 'Institutionalism: On the Need to Firm Up Notions of Social Structure and the Human Subject', *Journal of Economic Issues*, **37**(1): 175–207.

Lawson, T. (2012a), 'Ontology and the Study of Social Reality: Emergence, Organisation, Community, Power, Social Relations, Corporations, Artefacts and Money', *Cambridge Journal of Economics Narnia*, **36**(2): 345–385, doi: 10.1093/cje/ber050

Lawson, T. (2012b) *Reorienting Economics*. London: Taylor and Francis. doi: 10.4324/9780203929964

Lawson, T. (2016a) *Collective Practices and Norms*, Cham: Springer, pp. 249–277. doi: 10.1007/978-3-319-28439-2_11

Lawson, T. (2016b), 'Comparing Conceptions of Social Ontology: Emergent Social Entities and/or Institutional Facts?', *Journal for the Theory of Social Behaviour*, **46**(4): 359–399, doi: 10.1111/jtsb.12126

Lewis, P. A. (2005), 'Structure, Agency and Causality in Post-Revival Austrian Economics: Tensions and Resolutions', *Review of Political Economy*, **17**(2): 291–316, doi: 10.1080/09538250500067320

Loasby, B. J. (2001), 'Time, Knowledge and Evolutionary Dynamics: Why Connections Matter', *Journal of Evolutionary Economics*, **11**(4): 393–412.

MacKenzie, D. A. (2008), *An Engine, Not a Camera: How Financial Models Shape Markets*. Cambridge, MA; London: MIT Press.

Nonaka, I. (1994), 'A Dynamic Theory of Organizational Knowledge Creation', *Organization Science*, **5**(1): 14–37.

North, D. C. (1991), 'Institutions', *The Journal of Economic Perspectives. American Economic Association*, **5**(1), pp. 97–112. Available at: http://www.jstor.org/stable/1942704.

North, D. C. (2006), *Understanding the Process of Economic Change*. Princeton: Princeton University Press. Available at: https://books.google.com/books?hl=de&lr=&id=1RypQfxjQqEC&oi=fnd&pg=PR7&dq=north + douglas + understanding + the + history + of + economic + change&ots=nU-Jm0h6os&sig=SKIEGrunQuZCFYQKLCbELd7XiVk (Accessed: 11 August 2017).

Palley, T. (2017), 'A Theory of Economic Policy Lock-in and Lock-out via Hysteresis: Rethinking Economists' Approach to Economic Policy', *Economics: The Open-Access, Open-Assessment E-Journal*, **11**(2017–18): 1–18.

Polanyi, M. (1967), *The Tacit Dimension*. Garden City N.Y.: Anchor Books. Available at: http://www.worldcat.org/title/tacit-dimension/oclc/718091 (Accessed: 11 August 2017).

Rosenbaum, E. F. (2000), 'On Measuring Freedom', *Journal of Theoretical Politics*, **12**(2): 205–227, doi: 10.1177/0951692800012002004

Rosenbaum, E. F. (2001), 'Culture, Cognitive Models, and the Performance of Institutions in Transformation Countries', *Journal of Economic Issues*, **35**(4): 889–909, doi: 10.1080/00213624.2001.11506419

Runde, J. (2002), 'Information, Knowledge and Agency: The Information Theoretic Approach and the Austrians', *Review of Social Economy*, **60**(2): 183–208.

Searle, J. R. (1995), *The Construction of Social Reality*. New York: Free Press. Available at: https://ubbx5.bib-bvb.de/InfoGuideClient.upasis/singleHit.do?methodToCall=showHit&curPos=17&identifier=−1_FT_1040673375.

Searle, J. R. (2005), 'What is an Institution?', *Journal of Institutional Economics*, **1**(1): 1–22, doi: 10.1017/S1744137405000020

Searle, J. R. (2015), *Making the Social World*. Oxford: Oxford University Press. doi: 10.1093/acprof:osobl/9780195396171.001.0001

Sugden, R. (2015), 'On "Common-Sense Ontology": A Comment on the Paper by Frank Hindriks and Francesco Guala', *Journal of Institutional Economics*, **11**(3): 489–492, doi: 10.1017/S174413741500003X

Wobker, I., P. Kenning, M. Lehmann-Waffenschmidt and G. Gigerenzer (2014), 'What do Consumers Know About the Economy?', *Journal für Verbraucherschutz und Lebensmittelsicherheit*, **9**(3): 231–242, doi: 10.1007/s00003-014-0869-9

Wrenn, M. V. (2006), 'Agency and Mental Models in Heterodox Economics', *Journal of Economic Issues*, **40**(2): 483–491.

Cite this article: Rosenbaum E (2022). Mental models and institutional inertia. *Journal of Institutional Economics* **18**, 361–378. https://doi.org/10.1017/S174413742100059X