

# **Institutions, Mental Models and the Causes of Hysteresis in Economic Policy Making**

*Eckehard Rosenbaum PhD*

European Commission  
Joint Research Centre  
Via Enrico Fermi 2749  
21027 Ispra (Italy)

## Abstract

Hysteresis has been singled out as one of the root causes of policy lock-in and lock-out, thereby having far-reaching consequences for both the conceptualization and the implementation of economic policies by either locking-in or locking-out policy measures. The recent literature on the topic has identified a number of mechanisms through which hysteresis may emerge including network effects, increasing returns to scale or disequilibrium learning and thus shows that hysteresis is likely to be relatively widespread problem.

The paper examines the notion of hysteresis and suggests that it does indeed improve our theoretical understanding of economic policy making. However, the paper also identifies two, albeit interconnected, shortcomings of the debate. The first and arguably more fundamental problem is predicated on a flawed ontology of economic phenomena, which posits mistakenly, or so it will be argued, a fundamental distinction between, and therefore independence of, the socio-cultural domain and the economic domain. As a consequence - this is the second problem - economics sees itself primarily as an observer of the economy and much less so as an actor in its own right within the economy. Importantly, this rarely acknowledge role of economics does not primarily stem from the ubiquitous advisory function of economics (and economists) but from its transformatory role, i.e. its being a source of concepts, notions and tools which inform and thereby guide people's behaviour.

On this basis, the paper then proposes an extension of the theory of hysteresis by making use of the notion of mental models as thought processes about how something works in the real world. Accordingly a mental model is a representation of the surrounding world, the relationships between its various parts and a person's intuitive perception about his or her own acts and their consequences.

Importantly, mental model help to guide behaviour and thereby have social and economic consequences because they are shaped and modified by economics, be it in its academic or its popularized form. At the same time, mental model are also shaped by the outcomes and consequences of human action, thereby giving rise to a two-way relationship between human representation of reality as it is perceived and understood through reflection and interaction, and the very nature of this so-called reality, as an allegedly self-contained and independent entity. Due to this two-way relationship, however, mental models may trigger and sustain self-reinforced processes which, due to their persistence and resilience against disturbances, either lock-out or lock-in economic policy measures.

## I. Introduction

Hysteresis denotes the dependence of the state of a system on its history, i.e. on events that have happened at some point in the past or circumstances which have prevailed during some interval, and which have acted as “switch-on” or “switch-off” mechanisms. Hysteresis in economics is therefore the dependence of the state of the economic system on its history. For instance, hysteresis is often associated with long-term unemployment, the argument being that people remain unemployed even after the cause of their becoming unemployed, e.g. a demand shock, is no longer there. Possible reasons may be that the skills of the unemployed are outdated or that the long term unemployed exert only little pressure on wage setting so that wages remain too high (Blanchard and Summers 1987).

In a recent paper, Thomas Palley has singled out hysteresis as one of the root causes of policy lock-in and lock-out, arguing that the notion of hysteresis emphasises not only lasting effects of past events on the current state of the economy, but also brings to light an assumption of economic policy making which may not always be warranted (Palley 2017). Accordingly, prior policy changes are usually not seen as imposing restrictions on the set of future policy choices, thereby constraining the conceptualization and implementation of economic policies. On the contrary, policy changes are assumed to be reversible while the original set of policy options can always be restored if needed.

The foregoing assumption may turn out to be wrong, however, because, or so Palley argues, policies are liable to affect the political order and thus the conditions under which policies are formulated and decided upon. For instance, policies may affect the distribution of wealth and income and thus of the very endowments which given agents the power to prevent certain policies (Palley 2017). Palley then goes on to identify a number of both economic and political mechanisms through which hysteresis may emerge and thus purports to show that hysteresis is likely to be a relatively widespread phenomenon.

While these mechanisms are also liable to have a bearing on institutional inertia as a peculiar form of policy lock-in or lock-out, they do not capture, or so I shall argue, an important reason why institutions, and thus arguably the main focus of recent economic policies, are characterized by a high degree of stability (or lock-in). As I shall try to show, institutions are stable because of the way they are (re-) produced through social interaction. Against this background, I shall further surmise that institutional inertia is Janus-faced insofar as it is both a precondition for institutions to functions, and a factor which prevents or slows down institutional change.

Beyond identifying another reason for hysteresis, the paper therefore also seeks to contribute to the theory of institutions and institutional change by adding two arguments: first, a more refined distinction between constraining and facilitating institutions, and second, supporting the said interpretation of institutions as facilitators or enablers, a clarification of the role of mental models in this context.

The paper is structured as follows: in section II, I shall briefly summarize Palley’s recent contribution to the debate on hysteresis, complementing it by some remarks on hysteresis and institutions. Section **Error! Reference source not found.** discusses some relevant shortcomings of institutional theory and argues (Section IV) that the notion of mental model as suggested by Denzau and North (1994) is a useful step forward but needs to be complemented by a more comprehensive theory of social (inter-)action.

Section V examines hysteresis against the background of the foregoing discussion of institutions. Section VI concludes.

## **II. Hysteresis in economics and the role of institutions**

Palley (2017) has identified a number of both economic and political mechanisms responsible for hysteresis in economic policy making. In this section, I shall briefly describe the political mechanisms before developing a more general argument about hysteresis as institutional inertia. This is not to downplay the economic mechanisms but I take it that the focus of the economic debate has hitherto been on the latter at the detriment of the former. Thus some rebalancing may be justified.

The first mechanism put forward by Palley (2017) concerns exit costs. Accordingly, it may be much more costly to exit from a policy than to introduce the very same policy. Cases where such exist costs may matter are attempts to reverse privatisation or to break up (again) market concentration fostered by deregulation. As a consequence, previously implemented policies become locked-in. Another example concerns the introduction of the Euro whose replacement by a national currency appears go along with prohibitive costs, thereby making any exit from the Eurozone an extremely risky endeavour.

Another mechanism discussed by Palley (2017) concerns group power. Group power results from the fact that many economic policies have distributional effects and that agents, if confronted with the risk of losing previous gains, often form alliances with a view to defend (what they regard as) their acquired rights. Obamacare appears to be a case in point.

The third mechanism in Palley's account results from the fact that some policies may directly confer wealth and income on some economic agents and thereby provide them with endowments which can be used to influence the political process, be it via sponsoring certain politicians and their campaigns, be it via influencing public opinion through PR activities or even ownership of media channels.

The final mechanism suggested by Palley (2017) is what he has labelled Humpty-Dumpty hysteresis, meaning the destruction of social and organisation capital which, once destroyed, may be difficult to reproduce and then accumulate again. According to Palley (2017), the decline of union membership and the resulting difficulties of labour to defend its interest vis a vis capital could be mentioned here.

The foregoing examples suggest that economic policy is nowadays to a large extent about structural reforms. Although the term "structural reforms" is often used in a very loose way, comprising everything that is not typically conceived as Keynesian demand management, I take it that the understanding of structural reforms is usually synonymous with the engineering of institutional changes, i.e. changes to the regulatory framework of the economy ("the rules of the game").

Seen from this perspective, a large portion of economic policies nowadays concerns the introduction, modification or abolishment of institutions. Policy lock-in can then naturally be understood as institutional inertia, i.e. as the ability of institutions to withstand their change or abolishment. Policy lock-out, on the other hand, would comprise cases where the introduction of new institutions, or substantial modifications of exiting institutions, faces equally serious obstacles.

Without downplaying the importance of the mechanisms identified by Palley (2017), the above understanding suggests that further insights can be gained by placing the investigation of inertia firmly in the context of a general 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).

Needless to say that it would go way beyond the scope of a short paper to provide a comprehensive account of institutions and institutional change. What I try to achieve instead aims to be more modest. I intend to 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 understanding hysteresis. The first aspect concerns the different types of rules which make up institutions as social rules. In particular, I will argue that the widespread if not predominant understanding of rules as constraints to be enforced is misleading (but see Hodgson (2006) for some remarks that point into a similar direction). While for some institutions, enforcement is indeed an issue, there are reasons to suggest that important institutions are self-enforcing.

Secondly, I will argue that institutions as socially reproduced 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 the very bearers of institutions

### **III. Institutions as/and rules**

#### **A. Towards a typology of rules**

Hodgson (2006) defines institutions as systems of established and prevalent social rules that structure social interactions. Accordingly, language, money, law, systems of weights and measures, table manners, and firms (and other organizations) are thus all institutions in this sense. And so are of course the hobbyhorses of many institutional economists: property rights. The above list suggests that the content and form of institutions vary considerably. Some institutions are codified (money, law), others are not (table manners), or only partially so (language). Some institutions are legally binding (mostly law), others are only partially so (language) or not at all (table manners).<sup>1</sup>

In Douglas North's account, institutions are mostly seen as constraints on behaviour (North 2006), which need to be enforced by someone somehow. At the same time, he also claims that rules provide incentives for complying with them. It is not evident how to reconcile these two features of institutions

---

<sup>1</sup> Given that the terms "institutions" and "organisations" are often used interchangeably, not least in the context of the European Institutions (Council, Commission, Parliament etc.), recall also Hodgson's definition of organisations. Accordingly, "[o]rganizations are special institutions that involve (a) criteria to establish their boundaries and to distinguish their members from non-members, (b) principles of sovereignty concerning who is in charge, and (c) chains of command delineating responsibilities within the organization." (Hodgson 2006). From this perspective, the European Institutions are essentially organisations, and by virtue of being an organisation they are also institutions. Thus not every institution is also an organisation.

unless enforcement is seen as a kind negative incentive. But perhaps this contradiction is more apparent than real anyhow. So as Hodgson (2006) suggests, some institutions may also play an enabling role. That is, they allow us to do (or facilitate doing) certain things. But if this is the case, then it would seem rather odd to speak about constraints in the first place. North may therefore have had different types of rules in mind when he talked about constraints and incentives respectively. Let's look at the distinction in more detail (see Figure 1).

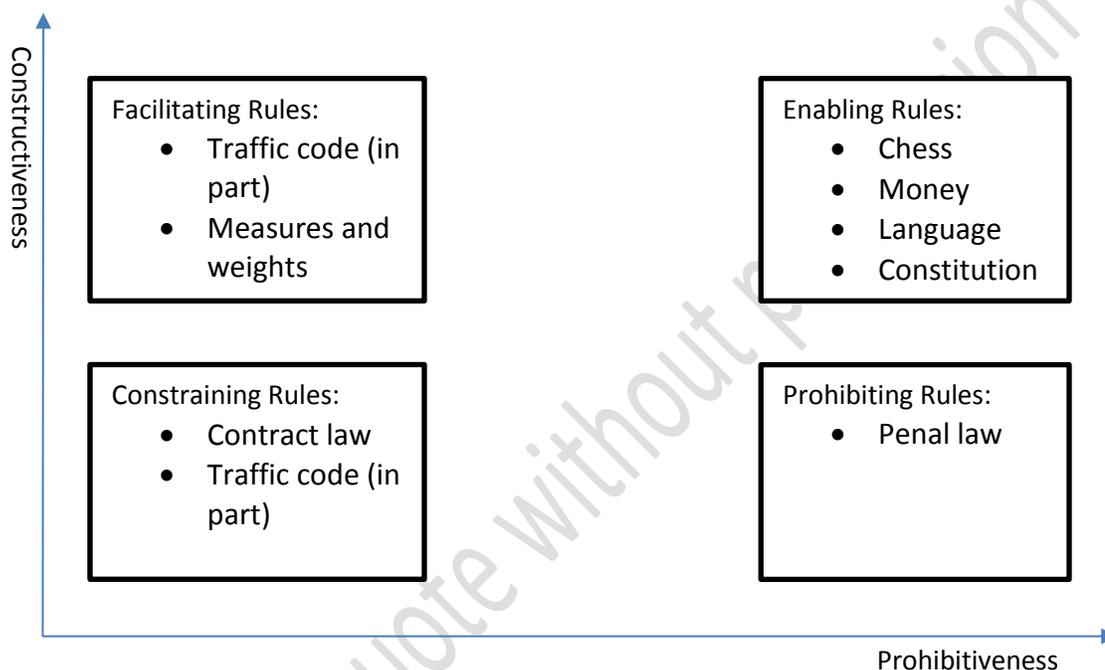


Figure 1: A Typology of Rules

“Thou shalt not kill”, the Seventh Commandment, is probably one of the most well-known, widely established 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 ...”) means that many societies know exceptions to that rule, be it self-defence, euthanasia or war. Despite the obvious benefits of that rule for society at large, 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 – which often translates into the question of whether punishment is “sufficiently deterrent” – is a different matter altogether and doesn’t need to be discussed here. 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 in the same sense as a prohibiting rule does exclude certain actions qua action type. Rather, a constraining rule can be said to reduce the range of admissible choices, where range is in turn conceptualised in terms of the (measurable) characteristics of an action or behaviour (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 resulting choice set is actually 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 unprohibited actions.

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. Also considerable parts of civil law can be summarised under the heading constraining rules in that these rules predominantly constrain my freedom of choice and action without doing so categorically for certain types of actions. 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 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 indeed their being constraints. They are what they are by virtue of what they constrain or prohibit. Let me now discuss two (again related) types of rules which 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.

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 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 to be made here is of course that it doesn't seem to matter where we drive provided drivers *agree* on the side. One can of course imagine a situation where this problem is negotiated between the drivers 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.

While also constraining behaviour, rules of the kind just discussed appear to achieve one thing above all others: they facilitate action by coordinating individual behaviour. That is why they will be referred to in what follows as *facilitating* rules. Facilitating rules, then, solve coordination problems by formulating solutions to coordination games of the kind described in the previous paragraph. More generally, they 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, which have been developed in the course of the scientific and industrial revolutions and which continue to be developed today.

The far-reaching division of labour which characterises virtually all production processes nowadays and which, in conjunction with technological progress, has led to enormous productivity gains, would have been impossible without common measures, norms and standards (North 2006). Amongst other things, they allow producers to produce for, and source from, large and relatively anonymous markets and thus to exploit to the fullest economies of scale and scope. None of that would be possible if the specifications of each and every product, including all its parts and components, would be up for negotiation and agreement each time a potential producer and buyer want to enter into a deal. Moreover, the exchange of information about a product or a service would become much more difficult and cumbersome without having the possibility to resort to a common technical language.

The advantages of common standards and norms are not confined to producers; they also exist for final consumers. Not only would it be rather difficult to compare the prices of final products if their properties had been measured in different units, all but the simplest commodities come along with a description of their technical properties (size, weight, compatibility with other products etc.), without which a consumer cannot even know whether the product has the sought after characteristics.

Couched in these terms, the issue of enforcement also looks distinctively different for facilitating rules. If confronted with the choice whether to comply or not, it almost goes without saying that compliance is the preferable option – not because non-compliance would be heavily sanctioned, but because compliance itself brings tangible benefits for the actor. A producer of printing paper for instance would take considerable risks by offering formats other than A3, A4 or A5 and so would a producer of printers which require non-standard size paper. By complying with the industry standards, both are on the safe side and do not have to worry about the compatibility of their product with related products.

Before moving to the next type of rule, it should be emphasised that the importance of coordinating rules is not due to their specific content (as the coexistence of metric and imperial measures illustrates), but to the fact that they are widespread and widely accepted and that their importance increases with their mutual acceptance and use.

The next type of rule which I would like to discuss can be characterised by a paradox. Enabling rules as I would like to call them 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 (albeit more difficult) 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 meaning, syntax and grammar would be impossible. That is why we do not understand a foreign language without learning all that and why we have to resort to another type of language (e.g. signs, gestures) which is simpler and more widely understood 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 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 a different language. 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 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 very constraints also ensure their functionality. Nevertheless, even enabling rules are to some extent arbitrary in the sense that even very different sets of rules can fulfil the same function(s), as the huge number of languages on Earth demonstrates.

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 and their customers and therefore the issuance of book money via credit is arguably one of the most important sets of enabling rules in the economic sphere. 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). But unlike the rules which enable the monetary system (those describing the role and functioning of central and private banks and more specifically the nature and content of balance sheets, accounts, various types of transactions, means of payment etc.), mostly constraining rules may often prove 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, however, many of them are not of intrinsic importance. They would not make much sense outside of the context of for instance the monetary system.

The difference between both types of rules can perhaps best be seen by considering the subset governing the issuance of paper money. 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 (although there are examples of parallel currencies).

## **B. Role, actor and status**

The typology of rules which has been developed above is incomplete. First of all, it is incomplete because an important type of institution has arguably not yet been discussed. Secondly, it 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. Thirdly, the relationship between rules and actors needs to be further clarified. As I will argue, all three 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, i.e. rules of the form "X counts as Y in context C".<sup>2</sup> 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.

---

<sup>2</sup> Some of the ideas reported here had already been developed in (Rosenbaum 1997)

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. This situation is unsatisfactory insofar as the four types of rules 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. The rules governing an organisation such as a firm address its management and its employees, and to some extent, its customers and suppliers, but not the employees of a firm in a different country. 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 employee and husband and car drivers, or they are the member of a sports club and a pensioner. As a consequence, multiple sets of institutions apply to them, not necessarily all of them all the time but at least with certain regularity. So far so good, 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 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 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 be undertaken irrespective of the properties of the object. So to use a sponge as a hammer would be rather difficult. But both an iron hammer and a big stone can be used as, and therefore assigned the function of, a hammer. So it is often not obvious from looking at an object what its precise function is likely to be.

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 summarize the above, 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 which 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 actors has a status which makes that set of rules applicable to the actor.

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 to treat these rules as institutions which are different in kind from Searle's. By contrast, enabling rules and constitutive rules appear to be similar to the extent that one might be tempted to regard them as synonymous. I shall argue 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. A show of hands in an assembly counts as vote. 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 is probably 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 of such a reformulation compared to a more straightforward formulation seems questionable. At least equally important, though, there are rules which describe how to play chess well, how to attack and how to defend. 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.

These considerations suggest that in particular enabling rules must comprise (or go hand in hand with) some notions on the part of the actors as to why and with what purpose a set of rules exists and is used. Actor must understand why they are doing what they are doing and how they can increase their chances of being successful.

## **IV. Institutions and mental models**

### **A. 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 (C. Lawson 1994; Searle 2005).

Denzau and North (1994) have introduced the concept of mental model, 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 section. Importantly, the facts which are

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.

Thus mental models embody the knowledge about how the world around us with which we interact functions and responds in turn to our actions. As a consequence, mental model guide decisions and choices by identifying the options which, as we believe, help us achieving our goals while excluding those which run counter to our objectives or at least do not further them. Thus they help to form expectations about the environment (Denzau and North 1994; Holland et al. 1986).

Several features of mental models should be noted. First of all, mental model 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).

Secondly, mental models embody explicit knowledge (as opposed to tacit knowledge (Polanyi 1967)). In other words, the knowledge which they comprise can be spelled out and communicated (and hence taught and learned). At the same time tacit knowledge is often complementary to the explicit knowledge contained in mental models (Nonaka 1994). 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.<sup>3</sup>

Thirdly, any mental model is necessarily partial in the sense that it does not provide a comprehensive account of reality (and does not intend to). 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).

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 which 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), which strategy to use and which to avoid.

---

<sup>3</sup> 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).

The view entertained here is that mental models provide the extra information and knowledge which is necessary in order to not only comply with, but to purposefully apply institutions. Mental models are transmitted culturally together with the institutions which they complement. As already pointed out, however, mental models are not static and cannot be so. Even chess players draw on the accumulated knowledge and experience of generations of chess players before them.

At this point, two complementary remarks appear to be in order. Firstly, the account of institutions given here does not see institutions as merely empirical patterns of behaviour as some old institutionalists have purportedly done (T. Lawson 2003), but institutions are seen as “social structures, irreducible to individuals, .. recognized as real, relational in nature and relatively enduring” (C. Lawson 1994). But these structures only exist because they are continuously reproduced through human (inter-)action. Thus “society is both the ever present condition and the continually reproduced outcome of human agency” (C. Lawson 1994) and institutions are hence an instance of the “duality of structure” emphasised by Giddens (2011). Importantly, there does not have to be an instance of the realization of these structure at any given moment in time for them to exist, it is sufficient that they are represented in people’s minds and thus can be drawn upon (or taken in account) once the occasion arises.

Secondly, and following from the foregoing remark, institutions cannot be seen as being independent of their mental representations as Denzau and North (1994) seem to insinuate occasionally. Thus society cannot, in a deep sense, be wrong about its institutions, only some individuals can. But when the mental models of an increasing number of people undergo changes which go into the same direction, then, inevitably, the institutions complemented by these mental models will also change because the way actors behave and thus reproduce these institutions will change.

Before moving on, I would like to make one additional point. In my view, the notion of reproduction in realist philosophy but also in Searle is somewhat vague. I 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 the issue, it seems that all three aspects are important for the reproduction of institutions as none of them appears to achieve that goal without the others.

## **B. Subject matter of mental models**

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 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 true economy looks like, only a very much simplified understanding is possible. But as pointed out above, such an understanding is not achieved via abstraction but inductively. Concomitantly, Austrian economists would emphasize that our

beliefs, even where accurate in some respects, are usually partial and often fragmentary (for example, 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, what they are and what we believe them to be is not independent. 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 (to some extent) it will. And if they believe instead that it functions according to a Keynesian model, then this belief too will leave an imprint on how the economy functions.

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 the functioning of financial markets, which, once internalised by economic actors into their mental models, guided their behaviour in ways consistent with the theory, they 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 held by theorists.

It is of course difficult to make generalisations about the content of people's mental models in general and their economic mental models in particular. 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.<sup>4</sup> So rather than asking about imputed causalities, the investigation of mental models requires, or so it seems, an approach which 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).

---

<sup>4</sup> 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 which raise ontological issues.

## V. Hysteresis and mental models

If institutions are based on (or rooted in) mental models, then their change presupposes a change of the corresponding mental models. Such a change is likely to be gradual, at least in the sense that not all holders of a specific mental model will simultaneously switch to another mental model.

This would seem to apply in particular to coordinating and enabling rules which are self-reinforcing and therefore depend to a lesser extent on the existence of an external enforcement mechanism (which would raise of course the question of who enforces enforcement).

Seen from this perspective, institutions are not only stable because they are useful (functional argument) but because they are rooted 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 inherently stable. Or, to put it differently, they will resist change and therefore exhibit inertia.

It should be emphasised though that institutional inertia so conceived is not *per se* a disadvantage or shortcoming of such institutions but a necessary condition for their functioning and their reproduction. A set of rules can only become (and remain) a socially accepted set of rules if (new) members of society can learn and familiarise themselves with these institutions and with the mental models that go hand in hand with them. However, such learning and habituation processes necessarily take time as does virtually any kind of learning that goes beyond mere imitation.

## VI. Conclusions

This essay has sought to make two points. First of all, it has argued that institutional inertia is bound to be an important reason for hysteresis and therefore policy lock-in in economic policy making. This is so because institutions as enabling socially accepted rules are not only self-reinforcing, they are also often complemented by complex culturally transmitted mental models. Both features prevent rapid changes. 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.

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 provide structure. Thus their stability is a precondition for their social acceptance and transformative reconstruction.

## VII. References

- Blanchard, Olivier J., and Lawrence H. Summers. 1987. "Hysteresis in Unemployment." *National Bureau of Economic Research Working Paper Series* 3(1): 288–95.
- Bucher-Koenen, Tabea, and Annamaria Lusardi. 2011. "Financial Literacy and Retirement Planning in Germany." *Journal of Pension Economics & Finance* 10(4): 565–84.
- Craik, Kenneth James William. 1967. *The Nature of Explanation*. Cambridge [etc.]: Cambridge University Press.
- Denzau, Arthur T., and Douglass C. North. 1994. "Shared Mental Models: Ideologies and Institutions." *Kyklos* 47(1): 3–31.
- Faulkner, Philip, and Jochen Runde. 2009. "On the Identity of Technological Objects and User Innovations in Function." *Academy of Management Review* 34(3): 442–62.
- . 2013. "Technological Objects, Social Positions, and the Transformational Model of Social Activity." *MIS Quarterly* 37(3): 803–18.
- Giddens, Anthony. 2011. *New Rules of Sociological Method a Positive Critique of Interpretative Sociologies*. Cambridge: Polity Press.
- Hodgson, Geoffrey M. 1988. *Economics and Institutions: A Manifesto for a Modern Institutional Economics*. Cambridge: Polity Press.
- . 2006. "What Are Institutions?" *Journal of Economic Issues* 40(1): 1–25.
- Holland, John H., Keith J. Holyoak, Richard E. Nisbett, and Paul R. Thagard. 1986. *Induction: Processes of Inference, Learning, and Discovery*. 5th print. Cambridge Mass. [u.a.]: MIT Press.
- Jappelli, Tullio. 2010. "Economic Literacy: An International Comparison." *The Economic Journal* 120(548).
- Johnson-Laird, P N. 2004. "The History of Mental Models." In *Psychology of Reasoning: Theoretical and Historical Perspectives*, ed. Ken I Manktelow and Chung, Man Cheung. Psychology Press, 179–212.
- Lawson, Clive. 1994. "The Transformational Model of Social Activity and Economic Analysis: A Reinterpretation of the Work of J.R. Commons." *Review of Political Economy* 6(2): 186–204.
- Lawson, Tony. 2003. "Institutionalism: On the Need to Firm up Notions of Social Structure and the Human Subject." *Journal of economic issues* 37(1): 175–207.
- Loasby, Brian J. 2001. "Time, Knowledge and Evolutionary Dynamics: Why Connections Matter." *Journal of Evolutionary Economics* 11(4): 393–412.
- MacKenzie, Donald A. 2008. *An Engine, Not a Camera : How Financial Models Shape Markets*. Cambridge, MA; London: MIT Press.
- Nonaka, Ikujiro. 1994. "A Dynamic Theory of Organizational Knowledge Creation." *Organization science* 5(1): 14–37.
- North, Douglas C. 2006. *Understanding the Process of Economic Change*. Princeton: Princeton University Press.
- Palley, Thomas. 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, Michael. 1967. *The Tacit Dimension*. Garden City N.Y.: Anchor Books.

Rosenbaum, Eckehard F. 2000. "On Measuring Freedom." *Journal of Theoretical Politics* 12(2): 205–27.

Rosenbaum, Eckehard F. 1997. "Some Remarks on the Ontology and Epistemology of Social Reality and Institutions." In *Proceedings of the XXII International Colloquium of Economic Psychology*, Valencia: Promolibro, 601–11.

———. 2001. "Culture, Cognitive Models, and the Performance of Institutions in Transformation Countries." *Journal of Economic Issues* 35(4): 889–909.

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

Searle, John R. 1995. *The Construction of Social Reality*. New York: Free Press.

———. 2005. "What Is an Institution?" *Journal of Institutional Economics* 1(1): 1–22.

Wobker, Inga, Peter Kenning, Marco Lehmann-Waffenschmidt, and Gerd Gigerenzer. 2014. "What Do Consumers Know about the Economy?" *Journal für Verbraucherschutz und Lebensmittelsicherheit* 9(3): 231–42.

Draft - Do not quote without permission