

# Rationality and Rule Following

## On Procedural and Consequential Interests of the Rule-guided Individual

Jukka, Kaisla

### *Document Version*

Final published version

### *Publication date:*

2001

### *License*

CC BY-NC-ND

### *Citation for published version (APA):*

Jukka, K. (2001). *Rationality and Rule Following: On Procedural and Consequential Interests of the Rule-guided Individual*. Institut for Industriøkonomi og Virksomhedsstrategi, Handelshøjskolen i København. Working Paper / Department of Industrial Economics and Strategy. Copenhagen Business School No. 2001-11

[Link to publication in CBS Research Portal](#)

### **General rights**

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

### **Take down policy**

If you believe that this document breaches copyright please contact us (research.lib@cbs.dk) providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 25. Jan. 2025



**Rationality and Rule Following:**  
On Procedural and Consequential Interests of the  
Rule-guided Individual

**Jukka Kaisla**

Department of Industrial Economics and Strategy  
Copenhagen Business School  
Howitzvej 60, 2000 Frederiksberg, Denmark  
[jukka.kaisla@get2net.dk](mailto:jukka.kaisla@get2net.dk)

May 2001

It is not the purposive but the rule-governed aspect of individual actions which integrates them into the order on which civilisation rests.

(Hayek 1978, 85)

## Introduction

The purpose of this paper is to examine connections between rationality and rule following, and to propose that a distinction between consequential and procedural interests can help us in explaining rule following behaviour. Simon (1955, 1978) maintains that when modelling rationality, more attention should be directed to the processes by which the actor arrives at choices — instead of preoccupying solely with the results of rational choice. Whatever methods the chooser is assumed to be using when rationally deciding upon an appropriate course of action, it appears that rules play a central role in such processes (e.g., search rules, threshold rules for defining satisficing levels, etc.). Limited or bounded rationality examined solely with respect to the consequences of rational choice remains as tautological as rationality in its omniscient version.

Hayek's (1952) theory of mind examines how the individual comes to perceive events in the first place. The general working properties of the mind and perception formation can clarify some central aspects of decision processes. Two ideas that become central to perception formation and decision making are: (1) perception is based on the *categorising* disposition of the mind and (2) perception formation requires constant *(re)interpretation* and interplay between accumulated categories and disclosing events.

The notions of categorisation and interpretation become interesting when rationality is examined in connection with rule following. A preoccupation with the consequences of choices disregards the fact that choices are arrived at by the rule-following disposition of the mind. On the other hand, rule following as a behavioural mode requires interpretation which may in some cases be very similar to the comparative assessment of the consequences, as examined in the rational choice model. An approach called *rule-individualism* (Rowe 1989, Vanberg 1994) aims at replacing 'the *logic of choice* by a genuine *theory of behaviour*' (Vanberg 1994, 7, emphasis in original). The present paper will maintain that rule-individualism remains preoccupied with the results of rational choice and therefore fails in its attempt to fulfil its goal.

As a partial remedy to this problem, the present paper proposes a distinction between consequential and procedural interests in choice behaviour. If we allow the actor's rational interest to be directed to the appropriateness of her action regarding the rules that she knows of, and to the interpretation of the behavioural recommendations of those rules in particular situations, her choice behaviour is no longer defined by the results of rational choice alone. Instead, choice behaviour is then better described by the procedural interest in finding an appropriate rule and interpreting its behavioural recommendation in a given situation. It should be noted that the distinction between procedural and

consequential interests in my treatment deviates from the idea that rule following is by definition based on procedural rationality whereas case-by-case adjustment is defined by consequential considerations. As will be explained in the paper, rule following may be motivated by consequential interest (as in rule-individualism) as well as by procedural one.

This paper is organised as follows: I start by discussing the notions of rationality and rule following. In the second section the central ideas of Hayek's theory of mind will be presented. Section three deals with rule-individualism together with some logical problems it comprises. In section four the distinction between procedural and consequential interests will be discussed. Finally, section five will provide some concluding remarks.

## Rationality and rule following

Rational choice theory (RTC) recommends forming explanations of social phenomena beginning with the individual's choice behaviour which is viewed as being rational. Rationality manifests itself in the purposeful action of the individual. The unit of analysis is a choice made by an individual. The theory suggests that an individual will always choose an alternative that maximises her utility (or minimises her disutility) in any given situation (cf. Coleman 1990, 13-9). What these choices are about does not concern RTC as it is essentially about how well means are applied in the pursuit of ends, rather than defining the aims of individuals (Elster 1986, 1). An action can be said to be rational if the individual has reason to believe that her chosen course of action is the best means to attain whatever she is aiming at.

As a general approximation rational choice is consequentially oriented in the sense that goodness of a choice is defined by the results it is expected to bring about. Also, rational choice is generally understood to concern case-by-case adjustment in separate choice situations. In the present use, case-by-case adjustment always refers to consequential considerations. By the term *procedural rationality* Simon is concerned with the lack of procedural consideration in rational choice. For Simon, procedural rationality means 'the effectiveness, in the light of human cognitive powers and limitations, of the procedures used to choose actions' (1978, 9). For epistemological reasons relating to limited reason and genuine uncertainty, the present approach does not maintain that we could, even in principle, demonstrate that rule following would be *apriori* a more effective behavioural mode than case-by-case adjustment — or vice versa. Quite the contrary, this paper will propose that the agent may be cognitively capable of switching between procedural and consequential considerations. Whether her interest is directed to either type is another issue.

By directing attention to processes by which the agent comes to choose actions, Simon opens the door to the realm of rules. Rules can then enter decision-making models in various ways. For instance, it allows functional arguments of the type: if the agent follows a certain rule in a certain type of recurrent situation, then rule following in that type of situation must be an efficient response. Such an argument may,

however, have to deal with the efficiency of habitual behaviour. A problem with habitual behaviour is that it is in many ways a complete opposite to the general idea of rational behaviour, as it is unresponsive to situational particularities and devoid of conscious reflection by the decision maker. It seems that we cannot distinguish rule following as a behavioural regularity from case-by-case adjustment without, to some extent, relaxing the requirement for situational judgement (Vanberg 1994, 33).

Another way for rules to enter the process of choosing is implied in Hayek's theory of mind (1952). All types of choice behaviour are based on the rule following disposition of the mind. Furthermore, all action, whether rule following or case-by-case adjustment requires interpretation. This implies that both rule following and case-by-case adjustment share the basic cognitive and behavioural requirements.

## Perception as a Process of Classification

Rational choice theory describes the choice process as one that begins at the recognition of available alternatives, continues by the evaluation of their respective expected consequences and ends at choosing the most preferable option. It is, however, unable to explain how individuals acquire and use knowledge to pursue rational choice behaviour in the first place. A central implication of Hayek's theory of mind (1952) is that *every* type of action, whether rule following or case-by-case adjustment, is based on perception formation through the categorising disposition of the mind. The working properties of the mind can thus be specified as rule following. But if the rule-following disposition of the mind does not discriminate between case-by-case adjustment and rule following at the observable action level, then we need to search for an explanation of rule following as a behavioural regularity elsewhere.

Central to Hayek's theory of mind is the notion of *interpretation*. This notion is also important for the present discussion because it functions as a bridge between case-by-case adjustment and rule following. The central message of *The Sensory Order* is that every type of action, including rule following, requires constant interpretation. In the previous section it was discussed that action, in order to qualify as rule following, needs to be unresponsive toward the particularities of the situation the agent finds herself in. On the other hand, the agent faces a problem of choosing which rule to follow at particular types of situations. She has to interpret the situation even before a rule-following type of behaviour can commence.

A problem with the idea of constant interpretation is that individuals seem to also follow rules which they are not conscious of. Rules do not necessarily exist in articulated forms (Hayek 1973), or even articulable forms (Hayek 1952). In such cases, interpretation becomes rather an innate process of the mind as the individual may remain unaware of any interpretative effort. This feature relates to an interpretation of rules as behavioural patterns or regularities of conduct (Hayek 1967, 66). If rules are viewed as observable recurrent patterns of

behaviour, the problem of constant interpretation does not arise. The only thing that counts then is the behaviour itself, not whether it is an outcome of unresponsiveness to the particularities of events or of some interpretive effort, or any combination of these two.

Hayek's theory suggests that what we can perceive are the *recurring patterns* of separate situations (1967, 23). What our mind is trying to figure out when we are faced with a new situation are elements that show some resemblance to those that we have experience of. We are trying to find possible connections between the elements of the situation we find ourselves in and the categories we have accumulated through experience.

The human mind is, however, limited in the sense that we cannot go through the innumerable particularities of a new situation and compare them separately as an automaton with our cumulated experience to find common elements. The human mind is not developed to consider every detail in separate situations. The disposition of perceiving regularities, even though it facilitates the development of knowledge about causal connections between regularities, hinders us from perceiving any situation in its full detail.

Pattern recognition is based on our ability to *classify* elements of events (Hayek 1952, s. 2.32–2.38). The ability to discern recurrent patterns arises from our ability to create categories of recurring elements in dissimilar events. The individual does not respond to separate situations as unique events (in absolute terms), but instead tries to classify their elements into certain types, based on the similarities that she can discern between the elements of the situation at hand and the categories accumulated by experience. Each perception is influenced by previous classifications. A new event is always perceived in association with the accumulated structure of elements with which it has something in common. If the elements of an event had no relation to any of the accumulated classes, the individual would remain unable to perceive the event in the first place. 'If sensory perception must be regarded as an act of classification, what we perceive can never be unique properties of individual objects but always only properties which the objects have in common with other objects' (Hayek 1952, 142).

What is assumed to happen during classification and re-classification processes is also important. Everything we perceive is related to previously accumulated classes. But also, any event contains the potential to create new and change existing classes. (Re)classification is thus a process where new events intertwine with existing categories. In order for the mind to be able to perceive order, the accumulated classes must modify perception of a new event more than the other way around. If this were not generally so, new events would continuously break down existing structure of classes and the individual would lose the ability to perceive order.

Another interesting feature in the categorising process is the feedback mechanism between the individual and her environment. Environment is generally seen as providing the feedback to which the individual then adjusts her behaviour. The individual's learning process is based on the method of trial and error (Hayek 1967, Popper 1972) where trials are hypotheses drawn upon experience and their selection is based

on partly *subjective* evaluation of their respective successes or failures to achieve what is aimed for. What separates the present approach to learning from an alternative interpretation of trial and error is that not only are the trials viewed as representing the individual's subjective conjectures about causal connections, but also that the disclosing consequences are interpreted by the individual, and as the individual can perceive reality only through her subjective understanding, the degree of success or failure remains partly a subjective matter as well. This interpretation may have slightly different implications than an interpretation according to which real events work as the *objective* selection mechanism, discriminating between success and failure irrespective of the individual's assessment.

Classification is not necessarily a simple and straight forward process, however. An event may consist of elements that belong to more than one class at a time and they may also on different occasions be assigned to different classes depending on the accompanying elements (Hayek 1952, 50). Classification may thus be *multiple* in these two separate ways. Furthermore, classification may take place in sequences across different levels of the hierarchy of classes. One classification act may in turn become a subject to further classification, and so on (*ibid.*, 51).

Hayek's theory of mind suggests that experience is essential for any formation of perception, that perception is essentially a process of classification of recurrent elements. The behavioural disposition of rule following is thus present in the very elementary processes by which we make the world intelligible to us.

## Rule-individualism

Rowe (1989) maintains that what he calls 'act-individualism', the behavioural description of rational choice theory, cannot explain socio-economic regularities, such as rules and institutions. 'If act-individualism were true, then social facts, social institutions, society, could not exist' (p. 4). This is because a self-interested maximiser would be unable to forego an opportunity to defect while others signal willingness to cooperate in mixed-motives situations presented in prisoner's dilemma.

Therefore, we will need an alternative behavioural theory to explain these regularities. For Rowe, the alternative is found in applying rational choice at the level of choices among *rules* of actions, instead of at the level of choices among actions themselves. 'A rule of action is rational if, by following that rule, an agent maximizes his expected utility' (Rowe 1989, 5). A single action cannot be judged rational as such, but only by considering to what extent it corresponds with a rule that is rational to follow. He concludes that 'social institutions are in fact nothing more than agents rationally following rules of action, and being believed by other agents to do so' (*ibid.*).

Rowe's rationale for rule following is based on an appealing idea for any rational choice theorist: if individuals are rational when buying and selling, then why should they not be rational in other activities, like

in choosing whether or not to follow a certain rule? He explains the basic logic of rational choice among rules as follows:

Whereas act-individualism proposes a one-step test of rationality, the action being evaluated directly in terms of its consequences, rule-individualism proposes a two-step test of rationality, the action being evaluated in terms of the rules to which it conforms, and the rule in turn being evaluated in terms of the consequences of following that rule (Rowe 1989, 23).

Therefore, Rowe concludes that:

[i]f the value to an agent of violating his rule exceeds the value to him of maintaining his reputation for following it, then he will violate that rule (Rowe 1989, 24)

Vanberg (1994, 31-2) notices that Rowe's rationality assumption is based on a kind of 'second order' case-by-case adjustment in the sense that the individual, instead of calculating which choice-option is rational to choose, calculates whether or not following a rule is rational. If violating a rule gives larger expected pay-offs than following that rule, then the individual will defect.

Rowe's view seems to introduce a logical problem:

- Postulate 1: an action is rational only in so far as it is part of a rational rule of action – it is neither rational nor irrational in itself (Rowe 1989, 5).
- Postulate 2: a rule of action is rational if, by following that rule, an agent maximizes his expected utility (ibid.).
- Hypothesis: if the value to an agent of violating his rule exceeds the value to him of maintaining his reputation for following it, then he will violate that rule (ibid., 24).

Insofar as postulate 1 holds, any action that is not part of a rational rule would be neither rational nor irrational. Violating a rational rule would then become neither rational nor irrational. As change in rules often requires a violation of some existing rule, a change in rules would then become neither rational nor irrational. If a change in rules becomes neither rational nor irrational, then rules themselves become neither rational nor irrational. Thus, the hypothesis seems inconsistent with postulate 1. The individual appears suddenly capable of defining rationality of an action that is not part of a rational rule.

Vanberg (1994) has also adopted the term 'rule-individualism' to define the behavioural foundations of the individual. The individual is unable to calculate the best course of action in separate, dissimilar situations and therefore adheres to mental processes which are not analysed in rational choice theory. She can use her past experience and her categorising ability to make conjectures about the consequences of her choice-options. Individuals are ascribed with 'the capability to learn from experience, and to adapt, over time, their repertoire of behavioural rules to relevant aspects of their environment' (ibid., 29).



By definition, the goodness of rules cannot be judged by their performance in a single situation. Rule-following means that the individual gives up the desire to evaluate every choice situation as a separate and that she commits herself to the rule that has worked well in the past. This notion is not necessarily shared by all advocates of rule-individualism, however. As above, e.g., Rowe (1989, 23) interprets that the meta-choice between case-by-case adjustment and rule following is a continuous case-by-case assessment process where the individual evaluates the potential outcomes of violating a rule against the past outcomes the rule has brought about.

For Vanberg, the essence of rule following is *not* to calculate in every choice situation, but, to some extent, to remain unresponsive to the changing particularities (Vanberg 1994, 33). To say that an individual *chooses* to follow a rule would, therefore, mean that the individual basically possessed the capacity to evaluate situations case by case, but would voluntarily give up her calculative capacity. As Vanberg puts it, 'she would have to decide, by rational choice, not to be rational' (ibid., 34). Thus, for Vanberg, the individual does not seem to possess a capacity to switch between rule following and case-by-case adjustment at will.

On the other hand, Vanberg views the rationale for rule following as being based on 'some comparison among potential alternative general patterns of behaviour' (1994, 17). To adopt a rule is then rational if it is expected to be more advantageous than an alternative strategy:

We can view an individual's adoption of a behavioural rule as being based on some comparison among potential alternative general patterns of behaviour. To adopt a rule in this sense can be considered 'rational' if it is found to be a more advantageous strategy than potential alternatives, where attempting to maximize on a case by case basis can be viewed as *one* alternative. ... In general it can be argued that adopting a rule for how to behave in certain types of situations is rational if rule-following can be expected to result in larger overall pay-offs (over a relevant period of time) than case by case adjustment. (Vanberg 1994, 17)

An important question arises about whether or not the individual is, even in principle, able to recognise that rule following will be on balance advantageous compared to case-by-case adjustment. The logic of reasoning that I am interested in here is as follows: if the rationale for rule following is based on our cognitive limitations that preclude case-by-case adjustment, then rule following describes choice behaviour in general. If, on the other hand, the individual is in fact able to pursue case-by-case adjustment but prefers to follow rules, then cognitive limitations alone do not provide the explanation for rule following.

To find out whether complying with a particular rule is more advantageous than rule-violation in the long-term may be difficult for the individual to establish. The individual needs to evaluate and compare potential consequences of both rule following and rule-violation in order to know whether or not the former is on balance more advantageous. This then would indicate that the individual's cognitive limitations do not explain rule following. A cognitive capacity is actually required in a

special sense to arrive at a rational choice to follow a rule. As the 'very nature of rules implies that their "goodness" can only be judged by their performance over a longer sequence of applications' (Vanberg 1994, 29), it becomes unclear how the individual, even in principle, could know when rule following is on balance more advantageous. This problem arises because if the individual has experience about following a rule, then she by necessity lacks the experience about the innumerable situations where she might have violated the rule. Any suggestion that she might know the latter cases (which have never been disclosed) fails to give a reasonable account of the fact that she does not even know the nature of the non-existent violations, that is, she does not know how precisely she might have chosen to violate the rule nor at what particular instant she might have done so, not to mention the possible consequences of doing so. The expected consequences of rule following can be viewed as being limitedly predictable, but only insofar as experiences from a rule have already been accumulated. But to claim that one can evaluate the consequences of future case-by-case adjustments or of those that might have taken place in the past would require mental capacities that are difficult to establish.

On the other hand, Vanberg does not view rule following as necessarily providing better overall consequences than case-by-case adjustment:

Following a rule rather than adjusting to the particular circumstances of each individual choice situation may involve a trade-off: the savings in decision making costs may have to be paid for by decreased overall 'quality' of choice-outcomes (Vanberg 1994, 18).

Three different types of rationales for rule following have been considered here: (1) cognitive limitations, (2) overall advantageous consequences of rule following and (3) savings in decision-making costs. The question whether or not the consequences of rule following can be viewed as being, on balance, more advantageous than case-by-case adjustments, according to a chosen criterion of goodness, is problematic. One can resort to a functional claim that if a rule exists, then it must be more desirable (and in that sense more advantageous). But this rationale has the same kind of irrefutability character as is found in rational choice theory. Savings in decision-making costs are an obvious consequence of reduced decision-making. An open question remains, however, about how we can balance these savings with the reduction of the quality of outcomes of non-existent activities. This refers to the fact that if the actor decides to follow a rule, then she foregoes case-by-case adjustment and therefore cannot know the quality of outcomes that the numerous separate choice situations would have resulted in if they had been chosen. It should perhaps also be noted that the explanation based on overall advantageousness of rule following and the cost-reduction explanation are potentially conflicting. The cost-reduction explanation suggests that case-by-case adjustment would in fact give more advantageous outcomes than rule following, whereas the advantageousness explanation claims the opposite. Both of these

explanations are at odds with the cognitive limitations explanation. If the individual is viewed as being incapable of case-by-case adjustment in the first place, then it is difficult to see how case-by-case adjustment could be viewed as an available behavioural mode to which rule following should be compared.

Cognitive limitations as a rationale for rule following introduces some interesting questions. A central question for the present paper is whether or not cognitive limitations discriminate between rule-following behaviour and case-by-case adjustment. This question arises insofar as the mental processes regarding both rule following and case-by-case adjustment are based on the same classification mechanism that is analysed by Hayek (1952).

### Procedural and consequential interests, and the cognitive capacity to switch between rule following and discretion.

There seem to be at least two interesting questions open here. The first question concerns our cognitive capacity. Vanberg criticises the view according to which the rationale for rule following could be based on calculation, either at the level of choices among actions or at the level of choices among rules of actions. Both rational choice theory and Rowe's version of rule-individualism are thus unsatisfactory in the light of the classification process advocated by Vanberg. Individuals follow rules precisely because they lack the capacity to evaluate separate situations in their full details. Even in a situation where the individual cannot find enough familiar elements to associate it with any already existing category, she uses the same experience-based classification process to establish a new tentative category. This is to say that when engaging in case-by-case adjustment, the individual is in fact using the same classifying process that rule following is based upon. Thus the classification act *per se* does not differentiate between rule following and case-by-case adjustment.

The second question introduces some new features to the above discussion as it asks whether individuals also follow rules in types of situations where their cognitive capacities would not prevent them from case-by-case adjustment. If the answer is in the affirmative, then cognitive limitations do not provide full explanation for rule following either.

Vanberg (1994, 33) views that rule following requires, to some extent, unresponsiveness toward the particularities of a situation. Unresponsiveness does not, however, necessitate the individual's *incapacity* of evaluating a situation. Whether or not an individual is capable of case-by-case adjustment and whether or not she uses this capability are two different questions. However, being able to evaluate a situation but refusing to do is not an available option for Vanberg. 'For a person to deliberately choose to follow a rule would require him/her to give up, by wilful choice, her capacity to calculate' (ibid., 33-4).

The present paper takes a different perspective to this issue. The distinction between procedural and consequential interests may help to clarify why a person may completely rationally choose to follow a rule retaining her capacity to calculate. If the individual's interest were assumed to be directed toward consequential assessment only, Vanberg's position would be justified. But, if we permit the individual to have interest directed to the procedural assessment, the picture changes. A central point to my position is that rule following requires *interpretation*. The individual needs to decide which rule to apply in a particular situation and how to interpret its meaning. If that choice and interpretation are directed towards consequential issues, that is, to figuring out which rule provides the best expected outcome, the individual is not giving up her ability to calculate, but is comparing the expected benefits that alternative rules would bring about. In procedural assessment, the benefits are not derived from expected consequences of alternative rules but, instead, from the expected appropriateness of a rule in a given situation together with the interpretation of its meaning in that situation.

Thus the individual does not give up her capacity to calculate; rather she is using that capacity motivated by the two types of interests, the procedural and the consequential. This perspective permits the possibility for the individual to switch between rule following and situational calculation in the sense that she is able to switch between procedural and consequential interests. This view fits well into the picture when we consider rule change. Changing rules requires some initial deviation or innovation. If the individual would only consider her procedural interests the whole time, rule change would occur as an unintended consequence due to the uncertainty of interpretation. Mistakes or variance in interpretation would give rise to new behavioural regularities. But insofar as the individual may in a situation where her action has previously been based on procedural consideration direct her interest toward alternative consequences, she may break the regular pattern and discover an entirely novel behavioural solution.

### **Procedural interest and learning**

None of the foregoing approaches seems to provide a completely satisfactory explanation for rule following. Rational choice theory does not view the individual as a rule-following actor to start with. Rule-individualism views choices among rules based on the same consequential assessment that is generally viewed as directing choices within rules. Categorising in Hayek's theory of mind explains the formation of perception as a categorising activity, but it does not differentiate between rule following and discretion at the observable behavioural level.

The aim of this section is to discuss how the individual may come to internalise procedural interest. It is proposed that insofar as rational contemplation is involved in the choice among rules, it may often owe more to the procedural than to consequential interests.

In his book *Ulysses and Sirens* (1979), Jon Elster suggests that individuals are not always fully rational. In the legend *Odyssey*, Ulysses, the king of Ithaca, has a potential dilemma during his journey. On the one hand, he would like to hear the call of the sirens, but on the other hand, he knows that nobody, after having heard their call, has been able to resist it and has thus been doomed to their spell for all eternity. Ulysses is aware of his limits of rationality and therefore designs a procedure that binds him (both literally and conceptually) to forego the undesirable action that otherwise would result in. He demands his crew to tie him to the mast and to block their ears so they are unable to hear his later commands.

Elster (p. 39-46) provides the principles of precommitment of this type as follows:

1. To bind oneself is to carry out a certain decision at time  $t_1$  in order to increase the probability that one will carry out another decision at time  $t_2$ .
2. If the act at the earlier time has the effect of inducing a change in the set of options that will be available at the later time, then this does not count as binding oneself if the new feasible set includes the old one.
3. The effect of carrying out the decision at  $t_1$  must be to set up some causal process in the external world.
4. The resistance against carrying out the decision at  $t_1$  must be smaller than the resistance that would have opposed the carrying out of the decision at  $t_2$  had the decision at  $t_1$  not intervened.
5. The act of binding oneself must be an act of commission, not of omission.

Principle 3 disregards types of decisions that do not have behavioural effects, like decisions to decide. According to principle 5, the fact that the individual prefers not to leave a given state is not viewed as evidence that she would freely have entered that state from all of the states that are open to her.

This type of precommitment can be viewed as rational within the framework suggesting that individuals are not fully rational. If individuals were fully rational precommitment would be unnecessary as the individual would be able to resist the later temptation to go against her 'true' preferences. Precommitment of this type is based on the consequential assessment of alternatives, however. The actor already knows what to expect to happen if she fails to precommit herself.

There may be another type of precommitment going on in the choice behaviour as well. It was suggested earlier in this paper that a central problem in evaluating different rules is that the individual often remains unable to assess the consequential goodness of alternative rules because rules are difficult to assess by reference to outcomes that do not yet exist. The evolutionary view on rules suggests that rule assessment is essentially retrospective. Only afterwards can we assess whether a rule produced types of outcomes that we prefer. But even then, we necessarily lack the knowledge of general outcomes of other rules that were available at the time we made the choice.

In the presence of ignorance about the comparative consequential efficiency of various behavioural alternatives the individual may direct her interest to the type of assessment that relates more to the

consideration of consistency of behaviour, that is, to the procedural assessment. Instead of asking what rule provides, on balance, the best average consequences, the individual may ask herself what would be the proper rule to apply in this situation, and how to interpret its behavioural recommendation. This is to say that a choice among rules that can be seen as rational contemplation may be based on the individual's interest in the procedural justification.

Due to epistemological problems concerning what we can know about rules and their outcomes, precommitment as a 'meta' rule is difficult to see as arising without social learning and interplay. Rawls' (1971) social-psychological construction provides for the possibility of a shared set of values and conventions to emerge. Rawls suggests that individuals who have a sense of themselves as individuals, and for whom pluralism with respect to final ends among all individuals is the rule, the only means to arrive at a social contract is through their sense of justice. If a contract is to have expected behavioural effect the individuals need to commit themselves to follow the agreed terms. But before individuals are willing to invest in a discourse leading to a potential contract they need to have expectations on reciprocity by others. By learning to precommit the individual establishes a quasi-stable reference point making her, to some extent, unwilling to defect even when defection would result in more desirable outcomes.

Chapter 8 of Rawls' *A Theory of Justice* explores how and under what conditions a sense of reciprocity arises from more primitive affections. The analysis builds upon psychological theories about stages in the child's development of moral attitudes. These theories suggest that sentiments of love and friendship, and the sense of justice itself, emerge from the experience of other people acting for our good. As a result of the learning process by which the child comes to recognise that others wish her well, she becomes precommitted to reciprocate in kind.

This type of precommitment counterbalances the suggested rational disposition to unilaterally defect while others cooperate in prisoner's dilemma situations. Precommitment as a quasi-stable reference point is tolerant towards experiences of defection by others. Precommitment is seen here as a more deeply rooted regularity than what can be considered a rational strategy in the reciprocal game of tit for tat (Axelrod 1984) which, as such, provides good reason to cooperate as well. Insofar as not all people defect all the time, precommitment may explain why people are willing to endure defection, contributing to a quasi-stable social order.

## Conclusions

It seems that a behavioural theory that limits its inquiry to the pure logic of choice cannot provide an explanation for the formation of social phenomena, such as rules, institutions and organisations. For instance, the mixed-motives game of prisoner's dilemma does not explain *per se* whether or not a cooperative pattern is reached and maintained. That is to say that the pay-off structure and the basic assumption that

individuals prefer better to worse do not suffice in providing an explanation of the emergence of a general behavioural pattern. As soon as other behavioural assumptions are introduced into the game, they seem to carry most of the explanatory burden. Assumptions of how much the players value e.g., the continuity of relations, trust, and reputation are pivotal to the outcome.

Rule-individualism seems to explain the individual's choice behaviour from the consequential perspective as well. The result is a second order rational choice among rules, or an emphasis on the cognitive limitations of the individual. But since we can observe that individuals engage in case-by-case adjustment as well, the cognitive limitations do not seem to explain rule following alone. Even though the individual's cognitive capacity is limited, she uses that capacity to develop expectations of the consequences that alternative choice options provide. This paper has argued that even though rule following may be partly motivated by consequential interest, the fact that the goodness of rules cannot be assessed as directly as is the case with separate choices within a framework of already existing rules directs the interest of the actor toward procedural considerations.

Efficiency claims for or against case-by-case adjustment or rule following are problematic. If we can observe one type of behaviour the other type of response will necessarily be missing. To say that due to our limited reason rule following is an efficient response to genuine uncertainty is problematic as individuals do engage in action that can be described as case-by-case adjustment. If case-by-case adjustment is refuted by reference to, for instance, Hayek's theory of mind which concludes that all kinds of action is based on the categorisation activity of the mind, then we can conclude that all types of action is rule following. But that would be a relabelling issue then.

When it comes to the issue of when it would be rational to follow rules instead of attempting case-by-case adjustment, things remain problematic as well. Some like to think that in a highly uncertain environment, rules need to be flexible or broad in order to provide room for proper adjustment to sudden situational changes. What this view implies is that actors in a highly volatile environment favour case-by-case adjustment over rule following. The problem with this view is that humans seem to behave in exactly the opposite way. In an increasingly uncertain environment we tend to delimit the range of behavioural alternatives (Heiner 1983, Dosi et al. 1999). We apply more rigid and narrow behavioural rules when things get volatile. The reason is rather obvious: if both the environmental factors and the range of possible response alternatives became unlimited, we would lose our means of orientation.

If efficiency here refers to proper response to environmental change, then both a broad and a limited rule may be treated as efficient. By increasing the number of possible adjustment alternatives a more flexible or broader rule would permit a 'correct' adjustment to take place, therefore qualifying the rule as efficient. And contrastingly, by limiting the range of possible response alternatives a more limited rule is an efficient response to the increase in environmental volatility. What is

efficiently eliminated is the risk of a response that might bring about harmful or fatal consequences. Both these views appeal to our immediate intuition. It can be tested whether or not people tend to resort to increasingly simple and clear rules in an increasingly complex and volatile environment.

Vanberg's version of rule-individualism appears beneficial in that it emphasises that a choice is always interconnected with the sequence of past choices. It argues that a choice is essentially a historical phenomenon, not something ahistorical and unconnected as viewed in the rational choice theory. The approach chosen here builds upon Vanberg's version on rule-individualism and suggests that a choice about which rule to follow is not limited to the consequential assessment. Insofar as a choice among rules requires interpretation, it is the interpretation based on the procedural interest that can explain which rule is applied in a particular situation.

The rationale for the conjecture about the presence of non-consequential interests emerges from the inherent inconsistency of rule-individualism. If the cognitive limitations explain rule following, then what explains case-by-case adjustment? If cognitive limitations do not explain rule following, and a choice among rules is based on rational expectations of the comparative consequences of both rule following and case-by-case adjustment, the assumption about human capacity becomes unrealistic. Neither version seems satisfactory.

A weakness of the procedural interests explanation might be that it is difficult to imagine choice behaviour that is not directed toward consequences of *some* sort. Searching for a proper rule and a proper interpretation of that rule in the particular context can be said to be consequential in the sense that the outcome that is aimed at is the correspondence between action and the rule that is considered proper. It might deserve mentioning that the notion of procedural interest does not aim at rejecting the idea of purposeful behaviour. The essential point remains that procedural interests describe preferences that give rise to a different type of choice behaviour than in consequential reasoning. A rational precommitment to long-term expectations at the cost of a short-term alternative is part of our daily decision-making. Empirical findings suggest, however, that human beings are equipped with a strong *status quo* preference that manifests itself at all levels of rules, from the personal to the social (Schlicht 1998). In her book *The March of Folly*, Barbara Tuchman examines historical incidents which brought about destruction, even though anyone with any sense at all would have easily been able to foresee the consequences. Procedural interests are consistent with *status quo* preferences but are not limited to them. The emphasis on the interpretative aspect of the search for a proper rule and its proper application may require mental exercise that exceeds the status quo preference explanation.



## References

- Axelrod, Robert (1984) *The Evolution of Co-operation*. New York: Basic Books.
- Coleman, James S. (1990) *Foundations of Social Theory*. Cambridge: Belknap.
- Dosi, G., L. Marengo, A. Bassanini, M. Valente (1999) Norms as Emergent Properties of Adaptive Learning: The Case of Economic Routines. *Journal of Evolutionary Economics*, 9, 5-26.
- Elster, Jon (1979) *Ulysses and the Sirens*. Cambridge: Cambridge University Press.
- Elster, Jon (1986) (ed.) *Rational Choice*. Oxford: Basil Blackwell.
- Hayek, F. A. (1952) *The Sensory Order. An Inquiry into the Foundations of Theoretical Psychology*. London: Routledge & Kegan.
- Hayek, F. A. (1967) *Studies in Philosophy, Politics and Economics*. London: Routledge & Kegan.
- Hayek, F. A. (1973) *Law, Legislation and Liberty, vol. 1, Rules and Order*. Chicago: The University of Chicago Press.
- Hayek, F. A. (1978) *New studies in Philosophy, Politics, Economics and the History of Ideas*. London: Routledge & Kegan.
- Heiner, Ronald (1983) The Origin of Predictable Behavior. *The American Economic Review*, 73 (4), 560-95.
- Popper, Karl R. (1972) *Objective Knowledge. An Evolutionary Approach*. Oxford: Clarendon Press.
- Rawls, John (1971) *A Theory of Justice*. Cambridge MA: Harvard University Press.
- Rowe, Nicholas (1989) *Rules and Institutions*. Ann Arbor: University of Michigan Press.
- Schlicht, Ekkehart (1998) *On Custom in the Economy*. Oxford: Clarendon Press.
- Simon, Herbert A. (1955) A Behavioural Model of Rational Choice. *Quarterly Journal of Economics*, 69 (1), 99-118.

Simon, Herbert A. (1978) Rationality as Process and as Product of Thought. *The American Economic Review*, 68 (2), 1-16.

Vanberg, Viktor (1994) *Rules and Choice in Economics*. London: Routledge.