Neurolaw is a decade-old interdisciplinary endeavour which aims at utilizing the findings of neuroscience in various legal disciplines. So far, the main focus of 'neurolawyers' has been on criminal responsibility as well as on the analysis of legal reasoning. However, I believe that much more may be gained through methodologically conscious reflection over the basic problems of law when they are seen in light of the developments in neuroscience and other biological disciplines. For example, it seems potentially fruitful to examine the fundamental presuppositions of the private and criminal law (e.g., that people are capable of acting in a purely rational way) against the view of human decision-making resulting from the work in neuroscience.

In this paper, I would like to go even further on the scale of generalization and abstraction. It is my intent to show that the current developments in biology (and in particular, in neuroscience and evolutionary theory) may be of use in the attempts to define law. The commonly acceptable definition of law has been for long the Holy Grail of legal philosophers. However, despite centuries of heated debates, no goal seems more distant than a firm answer to the question 'What is law?'

The question pertaining to the definition of law may be rephrased as the question 'What is a legal norm?', as there is a strong intuition that any definition of law must stay in some relation to the definition of a legal rule. Moreover, any theory of legal rules must address at least the following three issues. First: what is the...
ontological status of those rules or what kind of entity they are (the ontological question)? Second: what constitutes the legal rules' normative force or why they are and how they become objective reasons for action (the normativity question)? Third: where does the motivating force of legal rules come from or why people usually follow them (the psychological question)? Among the three, the normativity question is the most obscure. Let me analyse it in more detail. What does it mean that something is an objective reason for action? 'Objective reasons' are different than mere psychological motives (which are subjective). Moreover, if something is a reason then it might be used to justify a course of action, it may be weighed and compared with other reasons; put simply: reasons make it possible to reflect on our decisions as to how to act.

It is not to say that I believe the above characteristic to be an accurate description of the framework for addressing the problem of legal rules. In particular, it may turn out that the normativity question is ill-stated, i.e. that legal rules are never objective reasons for action. However, given relatively strong intuitions that stand behind them, and a long-standing legal-theoretic tradition, I take the three questions formulated in the preceding paragraph to be prima facie foundational to the problem of the definition of law.

I shall begin this paper with a detailed description of a theory of law developed some 100 years ago by Leon Petrażycki. It is interesting on its own, as one of the first conceptions that defines law in purely psychological terms. However, it will also serve to identify some key methodological problems, which must be dealt with in any attempt to 'naturalize law'. In the next section I shall sketch – against the background of Ludwig Wittgenstein's remarks on rule-following – an ontology of rules. In the following sections I shall try to justify this ontology by recourse to the findings of evolutionary theory and neuroscience. In turn, I shall reflect on the modularity paradigm in cognitive science, an approach to the study of the mind that seems, at least prima facie, to be incompatible with the conception of rules I defend. In the final section I shall apply my conception of rules to the problem of the definition of law; I shall also identify a number of methodological problems at the intersection of philosophy and neuroscience.
Leon Petrażycki is considered one of the most eminent – if not the most eminent – Polish legal theorist. His answer to the question of ‘what is law?’ is very original indeed. It does not mean, however, that his conception is flawless. Petrażycki set out to lay new foundations for jurisprudence and claimed that this task requires, first and foremost, an answer to the question of what is law. He observed that without such a definition any legal-philosophical and legal-theoretic considerations are carried out in a vacuum:

This is a principal and prejudicial issue, one which conditions the very possibility of the science of law.

The answer to the question ‘what is law’ demands, according to Petrażycki, the application of an adequate method. The problem is, however, that the methodological tools Petrażycki uses are out-dated, if not anachronistic – even from the perspective of the 19th century philosophy of science. Petrażycki claims that the goal of jurisprudence – understood as ‘a science in the correct sense of the word’ – is to look for the essence of law. In effect, one should try to construct a classical definition of law, i.e. a definition per genus et differentiam:

In order to divide the law scientifically into kinds and determine the differences between them, one should know the genus of law (…). It is thus necessary to recognize to which higher, more general category of phenomena law belongs.

Here, in a nutshell, one can find all the ingredients of the Aristotelian view of science. It assumes, first of all, that in the world there exist essences; the aim of science is, thus, to capture those essences in definitions which serve to build a

---


5 Cf. the views of French conventionalists.


7 Ibidem, p. 39.

8 Ibidem, p. 31, 35.
table of essential definitions (Porphyry’s tree), one that classifies all entities univocally. Petrażycki’s method presupposes, then, a very strong metaphysical view. It also encapsulates a static conception of science. Both those features are inconsistent with the practice of contemporary science.

Furthermore, one has to note that some of Petrażycki’s claims open the door to a revision of Aristotelian orthodoxy. He says, for example:

The foundation of scientific legal policy should be the examination of the causal features and the causal mechanisms of law in general, and of its different kinds and elements in particular.\(^9\)

The passage above indicates that jurisprudence should confine itself to the considerations of causal connections, while in the Aristotelian tradition the teleological connections play an eminent role. Moreover, Petrażycki also notes:

Especially in science, where each and every theory needs to overcome attempts at rejection and modification in order to be considered acceptable (...), where one has to do with a ‘struggle for life’ and only the fittest doctrines survive, one should expect that with the passage of time there should survive objectively sound theories.\(^{10}\)

Such a declaration fits well with the conceptions of Popper or Lakatos. This is evidence that Petrażycki had many original insights connected to the question of what is science. Unfortunately, he combined them with the anachronistic ideas of Aristotle. This led, in turn, to the development of an incoherent methodology. It wouldn’t be too devastating if the elements of the essentialist ideology remained at only the verbal level but Petrażycki applied those methodological rules meticulously, making his theory of law – one conceived in original sin – an unacceptable one.

For Petrażycki – and that is also Aristotelian – jurisprudence, as well as other humanistic and social disciplines, needs a foundation, which is to be found in a more basic science: psychology. Petrażycki was unhappy with the psychology of his time, in particular, disagreeing with the Kantian heritage it accepted. Kant divided mental phenomena into three categories: knowing, feeling and willing. Petrażycki considered this division incomplete, claiming that one should add to it another, fourth category:

One should distinguish not three, but four basic forms of inner experiences and four classes of

\(^{10}\) Ibidem, p. 205-206.
mental elements: (1) emotions, i.e. impulsions (two-sided mental experiences), 2) and 3) sensations and feelings (passive one-sided experiences), 4) processes of the will (one-sided active experiences).\(^{11}\)

Emotions (impulsions), a category which is essential for the task of defining law, are two-sided, active-passive, while all the other experiences are one-sided. Among the emotions Petrażycki includes hunger and love.\(^{12}\)

Furthermore, Petrażycki provides us with a classification of emotions. From our perspective, the most interesting are the ethical emotions (emotions of duty), which “are experienced as an inner limitation of freedom.”\(^{13}\) Ethical emotions could be further divided into moral and legal; the former are exclusively imperative (i.e., the actor feels obliged to do something), while the latter are imperative-attributive (the actor feels obliged to do something, but also accepts that someone else has the right to require her to do it). This distinction leads Petrażycki to the famous definition:

Law, as a separate class of real phenomena should be understood as such mental experiences whose emotions are of the attributive character (…). All the other ethical experiences, i.e. experiences of exclusively imperative emotions, should be deemed moral phenomena.\(^{14}\)

One should be mistaken, however, if she insisted that Petrażycki identifies legal norms with certain emotions or complex mental experiences. In order to clarify this issue, one should distinguish between such notions as emotion, representation, norm and duty.

Emotion, as noted above, is one of the four basic, irreducible mental experiences. People, according to Petrażycki, also have the power to imagine certain situations or behaviour. Such an imagined representation – together with the emotion it causes – constitutes a motive for action. In On the motives for action and on the essence of morality and law Petrażycki says:

We are interested, especially, in one particular type of motivation, the one in which there is a connection of representations of various acts with very peculiar emotions, which we deem

---

12 Cf. ibidem, p. 10.
13 Ibidem, p. 27.
ethical emotions or emotions of duty.\textsuperscript{15}

For instance, if I imagined taking part in a fraud, “I would experience a mental state similar to that which I experience while considering eating a piece of rotten meat, touching a spider or a snake; in normal circumstances I would experience repulsive emotions.”\textsuperscript{16}

Furthermore, one should distinguish between emotions, representations and motives on the one hand, and legal and moral norms and duties on the other. Petrażycki defines legal (moral) norms as the contents of the ethical (moral or legal) convictions.\textsuperscript{17} Thus, norms are not emotions or motives but they are of an intellectual character. They may be described as specific representations or propositions, which can be grasped (contemplated). It is worth noting that 'representation' (or 'proposition') serves as a primitive term in Petrażycki's psychology, similar to the notion of 'idea' in Hobbes, Descartes or Locke, or the notion of 'concept' in Kant.

The notion of duty is defined in a similar way. Petrażycki says:

Duties are ideal projections, which originate in our minds. Such projections are connected to the described emotions and representations, and not to some things or phenomena in the outer world.\textsuperscript{18}

It seems, therefore, that duties should also be called representations or propositions, which are graspable by the human mind. One can find the confirmation of those conceptual distinctions in the following passages:

The explained difference of genus between one-sided imperative (moral) and two-sided duty-imposing (legal) norms and duties is based on the adequate genus differences among the emotional-intellectual complex phenomena which are, as we demonstrated, the real base for ethical duties and norms.\textsuperscript{19}

The basic motivation, which consists in connecting representations of actions with the above characterized repulsive or impulsive emotions, we should deem ethical motivation and the corresponding principles of behaviour – ethical principles or norms.\textsuperscript{20}

\textsuperscript{15} L. Petrażycki, O pobudkach…, op. cit., p. 25.
\textsuperscript{16} Ibidem, p. 21.
\textsuperscript{17} Ibidem, p. 33.
\textsuperscript{18} Ibidem, p. 34.
\textsuperscript{19} Ibidem, p. 49.
\textsuperscript{20} Ibidem, p. 28.
Therefore, law and morality exist in the minds of people. Law cannot be identified with norms. It would be a mistake, however, to identify it with certain ethical emotions. When Petrażycki says that “law, as a separate class of real phenomena should be understood as such mental experiences whose emotions are of the attributive character”, he claims that law is a complex mental phenomenon, one that consists of adequate emotions, norms and duties.

Still, one more problem should be addressed: what is the relationship between the law (in Petrażycki’s sense) and the provisions of legal acts? Petrażycki’s reply is the following:

The representations of legal provisions or biblical commandments shall be deemed the representations of “normative facts”. Ethical convictions, to which such representations belong, shall be called positive ethical convictions, and their contents – positive norms. Ethical convictions, which lack such representations of normative facts, are intuitive ethical convictions, and the corresponding norms – intuitive norms.21

The general mental mechanism proposed by Petrażycki appears as follows: people have the capacity to imagine certain situations, patterns of behaviour etc. There are a plethora of sources of such imagined representations: legal acts, the Bible, or any other “normative fact”, as well as one's own intuition. Those representations cause the corresponding legal or moral emotion. Together, they serve as motives for action.

In light of the above, one may say that Petrażycki presents us with a peculiar ontology of law. He believes that norms are certain representations or propositions. He claims, moreover, that law cannot be identified with the set of legal norms. Law consists of complex mental states, which include representations (propositions), and emotions, together generating motives for action.

It is interesting that Petrażycki does not address the problem of law's normativity. He does not consider legal (or moral) norms as objective reasons for action. He seems to concentrate on a different question: how does a legal or a moral norm motivate people’s actions (i.e., the psychological question)? The key role in this process is played by the relationship between representations of certain

21 Ibidem, p. 33.
states of affairs and the ethical emotions they generate. Does this mean that the reductionist strategy deployed by Petrażycki leads to the elimination of the concept of law's normativity? Does Petrażycki show that the notion of a legal norm which is an objective reason for action is meaningless? The answer is a plain and resounding 'no'. The notion of normativity is needed as soon as Petrażycki moves from describing law to the problems of legal policy.

The entire theoretical enterprise described above, which aims at uncovering the psychological mechanisms of how legal and moral norms influence human behaviour, has an additional, practical goal. Petrażycki says:

The essence of the legal policy problems boils down to scientifically justified prediction of the effects of enacting legal provisions. Legal policy aims at developing such principles, which – introduced into the legal system or in some other way – would yield the required effects. Moreover, Petrażycki devotes much attention to describing those “required effects”. The legislator, he claims, has a certain goal to realize. The goal is to be reached in a purely instrumental way, with the utilization of the knowledge concerning the mental motivational mechanisms. In such a setting the legislator becomes a “super-human”. While “ordinary” people are led by emotions, the legislator applies the rules of pure instrumental rationality. It is necessary, of course, to explain the normativity of those rules. Furthermore, the surprising fact that a certain kind of rules – the rules of rationality – are of a totally different character to legal and moral rules is also in need of an explanation. Thus, the reduction proposed by Petrażycki is only partial and suffers from a serious case of 'schizophrenia'.

In conclusion: Petrażycki offers an intriguing conception of law. Contrary to usual presentations, he does not identify legal norms with emotions. He claims that law is a complex mental phenomenon. Unfortunately, along the way he commits some grave errors. The anachronistic method he applies – Aristotle's essentialism – carries with itself a serious metaphysical baggage. Accepting it, Petrażycki is forced to look for the essence of law, an ephemeral entity that is nowhere to be found. Moreover, misled by Aristotelianism, Petrażycki looks for a foundational answer to

the question 'what is law'. This foundationalism has two faces: firstly, it requires us to look for a science that is more basic than jurisprudence; secondly, it launches a search for some basic phenomena which, taken together, 'produce' law.

All those deficiencies would be avoidable if Petrażycki recognized the role philosophy should play in any attempt at defining law. Instead, he tried to construct the definition of law directly at the level of the “more fundamental” science of psychology (I disregard the fact that Petrażycki’s psychology has little to do with contemporary psychology, as my remarks are methodological in character). The disregard for the philosophical dimension of his project leads Petrażycki to a premature dismissal of the phenomenon of normativity. It turns out that the normativity problem – reduced to the psychological problem of the motivating force of law – reappears at a different level. When considering the goals of legal policy, Petrażycki explicitly states that a legislator should act according to rules of instrumental rationality, ones that clearly possess some normative force. The failure of Petrażycki’s theory vis a vis the puzzle of normativity may thus be described in terms of the unacceptable narrowing of the normative horizon. It is as if Petrażycki were content with the reduction of law's normativity to the normativity of the rules of instrumental rationality. Moreover, he seems to have no interest in explaining the normative force of the latter. In this way, the normative horizon he adopts is quite narrow: his theory aims at accounting for the alleged normativity of only legal rules. This may be interesting per se, but as a way to solve the normativity puzzle it is useless.

The failure of Petrażycki’s theory, apart from taking advantage of the anachronistic methodology, lies in the narrowing down of what I called 'the normative horizon'. In other words, his example shows that it is impossible to

---

account for legal rules and their normativity by concentrating exclusively on law. One needs to explain legal phenomena against a more comprehensive background, one that includes other kinds of rules and other kinds of normativity. Thus, we need to reflect on the nature of rules in general.

In *Philosophical Investigations* Ludwig Wittgenstein offers what is probably the most detailed analysis of the phenomenon of rule-following available.\(^{25}\) Let us consider Wittgenstein's favourite example. Imagine that you ask someone to add 2, starting with 0. In reply, you get the following sequence of numbers: 2, 4, 6, 8, 10, ... 1000. At this point, something surprising happens. After 1000 you hear 1004, then 1008, then 1012, etc. At your insistence that this sequence is incorrect – it should be 1002, 1004, 1006 – your interlocutor says that she is certain of her response. The question Wittgenstein poses in this context is: how do we know that ‘1002’ is the correct answer? He notes (§186):

(…) How is it decided what is the right step to take at any particular stage? — “The right step is the one that accords with the order — as it was meant” — So when you gave the order ‘+2’ you meant that he was to write 1002 after 1000 — and did you also mean that he should write 1868 after 1866, and 100036 after 100034, and so on — an infinite number of such propositions? — "No: what I meant was, that he should write the next but one number after every number that he wrote; and from this all those propositions follow in turn." — But that is just what is in question: what, at any stage, does follow from that sentence (…).

Wittgenstein’s question may seem crazy. That – in adding 2 – 1002 follows 1000 is obvious. However, it is the task of a philosopher to question the obvious. It is relatively easy to show the difficulty involved in capturing why a rule contains “all those propositions”. From the point of view of mathematics, the rule ‘plus 2’ is a function, i.e. an infinite set of pairs of numbers, of which the antecedent is the one to which we add 2, and the consequent is the result of addition. To know such a rule means to be able to contemplate the actual infinity. This we can't do, however. There must be a different way in which rules enable us to 'capture' the potentially infinite cases of their application. Therefore, we can note here an important intuition concerning the concept of a rule: rules are patterns of behaviour (or: contain patterns of behaviour). Let us call this requirement the **pattern condition**.

The second Wittgensteinian insight concerning the concept of a rule reads: rules are objective reasons for action (*the reason condition*). Wittgenstein notes (§222):

"The line intimates to me the way I am to go." — But that is of course only a picture. And if I judged that it intimated this or that as it were irresponsibly, I should not say that I was obeying it like a rule.

Put differently: even if we understood, how a rule contains “all those propositions”, it would not be sufficient to explain rule-following. Rules (at least some rules) must be objective reasons for action. Even if there are rules, how is it possible, that we *should* follow them? A complete theory of rules must reply to this question.

Against this background, Wittgenstein rejects some traditional conceptions of rules. The typical interpretation of *Philosophical Investigations* posits that Wittgenstein undermines three theories of rules: mechanicism, Platonism and mentalism. Firstly, rules are not dispositions to act in the given way (e.g., to answer ‘1002’). We do not follow rules 'automatically' and such an account is at odds with the reason condition. Secondly, rules are not mental states: Wittgenstein observes that one can imagine rules as some kind of picture. However, it is difficult to comprehend how such a picture can 'contain' all the cases of a rule’s application (e.g., for ‘plus 2’ this is an infinite number of cases). Thus, rules-as-pictures cannot account for the pattern condition. Thirdly, the same holds true for Platonism. Were rules platonic objects, we would be able to imagine them only as some kind of pictures.

I believe that the catalogue of the 'erroneous conceptions of rules' sketched above is too generous. In my opinion the correct way of interpreting Wittgenstein’s remarks is to observe that they are not uttered within an academic discussion, which clearly distinguishes between mentalism, Platonism and so on. The questions Wittgenstein asks are more concrete: what is a rule if it were to 'contain' all the possible instances of a rule's application? How can it be grasped? It seems that Wittgenstein's attack goes deeper than the rejection of mentalism or Platonism. His target seems to be the model of thinking-as-seeing. This is one of the most deeply rooted metaphors of Western culture. Thinking is usually referred to as seeing through the mind's eye. This idea had many progenitors but Descartes
played a leading role since it was he who urged us to look for ideas that we see in a 'clear and distinct' way. It was Descartes and his contemporaries who insisted that what matters is the mental discourse in our minds, the chain of ideas we see with our inner eyes, and not the 'public discourse', which is of no real philosophical significance.²⁶

Doubtless, the model of thinking-as-seeing is only a model, a handy metaphor that helps us to describe the process of thinking. Moreover, it seems useful in many contexts. When I try to solve a geometrical problem, recall someone's face, or imagine my lecture tomorrow, my thinking may resemble seeing. It is less useful, however, when I feel pain (Descartes claims that this also is an instance of seeing an idea) or follow a rule: here, the metaphor of thinking-as-seeing fails. From this perspective, Wittgenstein's remarks on rule-following undermine a key model of thinking and of the architecture of the mind, a model so dear to the Western philosophy. It is also at the same time an attack on mentalism and Platonism, to the extent that they are connected to the model of thinking-as-seeing (which they are, at least in their most typical incarnations).

It is much more difficult to present Wittgenstein's solution to the problem of rule following in a coherent way. However, a number of aspects of the solution may be identified. Firstly, Wittgenstein claims that one cannot speak of rule-following outside of a community. As Saul Kripke observes, “if one person is considered in isolation, the notion of a rule as guiding the person who adopts it can have no substantive content.” The situation changes when we “widen our gaze from consideration of the rule follower alone and allow ourselves to consider him as interacting with a wider community. Others will then have justification conditions for attributing correct or incorrect rule following to the subject.”²⁷

This observation is based on a number of Wittgenstein's remarks in which he stresses the communal character of rule-following. For example, he says:

The phenomenon of language rests on regularity, on agreement in acting. Here it is of the greatest importance that all of us, or the overwhelming number, agree on certain things. For example, I can be sure that the colour of this object will be called 'green' by most people

who see it. Thus, rules are neither mental states, nor platonic objects. One can speak of rule-following for we participate in a certain social practice; a practice based on common agreement.

Secondly, the sole regularities of social behaviour are insufficient to justify speaking of rule following. As Susan Hurley observes:

[I]t is not an adequate answer to say that the solution practices provide is a sceptical one, that nothing underwrites content and we just, contingently, happen to agree in doing this rather than that. (…) [T]he full force of the sceptical view dissolves our capacities for intentional action, for trying and choice, however arbitrary, as much as for perception and thought. It takes the ground out from under the feet of pragmatism and conventionalism, as much as Platonism and psychologism. It rules out appeals by the sceptic to our intentional responses, our attributions, our constructions, our investigations, our procedures of verification or ratification, etc.

In short: Hurley tries to say that in order to speak of rule-following one needs not only agreement in social reactions but also certain mental attitudes: that I believe that the given behaviour is obligatory (complies with a rule).

Thirdly, Wittgenstein claims that rule-following has a systematic character. Put differently, speaking of rule-following is possible only against a background of a complex system of behaviour and mental attitudes. One can say that one followed a rule only because there are a large number of cases in which one can speak of following different rules. Wittgenstein notes:

We say that, in order to communicate, people must agree with one another about the meanings of words. But the criterion for this agreement is not just agreement with reference to definitions, e.g., ostensive definitions — but also an agreement in judgements. It is essential for communication that we agree in a large number of judgements.

Fourthly, in light of the above, an important if not decisive role is played by very simple, basic rules. In §219 Wittgenstein observes:

"All the steps are really already taken" means: I no longer have any choice. The rule, once stamped with a particular meaning, traces the lines along which it is to be followed through the whole of space. – But if something of this sort really were the case, how would it help?

No; my description only made sense if it was to be understood symbolically. – I should have said: This is how it strikes me. When I obey a rule, I do not choose. I obey the rule blindly. Crispin Wright claims that most such basic cases of rule-following, in which one tends to speak of following a rule 'blindly', are the key to understanding the phenomenon of rule-following. These 'reactions' – although they are full-blooded rule-following, as “they are rational in the sense that they involve intentionality and a willingness to accept correction in the light of inadvertent breaches of the rule”31 – are to a great extent derived from our biological underpinning. Thus, it is no surprise that we agree when we determine colours: “[T]hat is no miracle: it is just what is to be expected of biologically and neurophysiologically similar creatures, hardwired and trained in similar ways.”32 On the other hand, the agreement in the most basic cases constitutes the foundation for the phenomenon of rule-following in more complicated situations, when it is more proper to say that following a rule is not 'blind', that it requires a 'decision' (cf. §186 of *Philosophical Investigations*). Thus, Wittgenstein observes that 'obeying a rule' has many faces (§235): there are cases in which we are prone to say that we follow a rule 'blindly' or 'unconsciously'; in other instances, however, obeying a rule is best described as a decision, when we consciously reflect upon the required course of action. Wittgenstein's key point is that the 'conscious' or 'complex' rule-following is possible because there is regularity or 'an agreement in action' between us, which is a manifestation of obeying 'simple' rules in an 'automatic' way.

Fifthly, at the basic, 'unconscious' level, the phenomenon of rule-following displays a feature that may be called 'relative normative unity'. What I have in mind is that at the basic level one cannot unequivocally distinguish between different kinds of rules. Some commentators claim that Wittgenstein speaks only of language rules yet this is a mistake. Nowhere does Wittgenstein narrow – at least explicitly – his remarks to language. Moreover, his favourite example – of the rule 'plus 2' – is hardly an instance of a purely linguistic rule. If it were, the same would hold of *any* rule: 'One should not kill other people', 'One should pay taxes', or 'One should

---

32 C. Wright, op. cit, p. 31.
maximize her utility function.' Of course, *Philosophical Investigations* are centred around language but Wittgenstein offers a new conception of the relationship between language and reality. It is no longer understood in terms of the language of describing or depicting the world. For Wittgenstein, language is a tool for *acting* in the world. In this way, the dichotomy of 'language − reality' loses its importance. One cannot clearly distinguish between linguistic and other kinds of rules: each rule, be it mathematical, moral or legal, has its linguistic dimension, is an element of a language-game, i.e. determines the uses of certain expressions. This is especially visible at the level of the most basic rules. Only through an *ex post* reflection, through the application of abstract categories of mathematical or linguistic rules, can we judge that the utterance of '1002' after '1000' was an instance of the application of a mathematical rule, while the utterance of 'green' was a result of applying a linguistic one.

One can justifiably ask how this Wittgensteinian conception fares vis a vis the pattern and reason conditions. The answer is straightforward, but subtle. The pattern condition – taken literally – is not fulfilled. However, in contrast to other theories of rules, there exists a rudimentary form of the fulfilment of the pattern condition in the Wittgensteinian account. On the one hand, it is provided by the regularities in social behaviour and related reactions to others' actions. On the other hand, people are equipped with what biologists and psychologists call a 'pattern recognition mechanism', i.e. a biologically conditioned capacity to grasp structural similarities between events, individuals or properties. The role of these abilities is clearly visible in the cases of the most basic rules. The reason condition in the developed conception is also present in a rudimentary form. Rule-following is a communal enterprise. In this way, it is not dependent only on what goes in our heads: we have (publicly shared) reasons to act in the given way, not merely psychological motives.

My final observation concerning Wittgenstein’s insights is this: Wittgenstein would never say that the *existence* of rules is conditioned by both the regularities of social reactions and mental attitudes. He would be prevented from doing so by his metaphilosophy, which forbids constructing philosophical theories, including
ontologies. Thus, to be faithful to the Wittgensteinian directives, one would need to say that the regularities of social reactions and mental attitudes condition the possibility of speaking of rule-following. However, as I have argued elsewhere, Wittgenstein's metaphilosophy is objectionable. In particular, it is based on some assumptions (e.g., regarding the possibility of distinguishing scientific or logical problems from the philosophical ones) that can – and should – be rejected.

Therefore, I am inclined to sketch the following ontology of rules which is inspired by Wittgenstein's remarks. At the most general level, rules may be described as entities which are conditioned by, or *supervene* on both mental attitudes and social interactions. More precisely, one can speak of two kinds of rules. First, the rudimentary rules emerge from the regularities of social behaviour coupled with special mental states. As I have tried to stress, rudimentary rules have certain specific characteristics. At this level, there are no individual, stand-alone rules. We must rather speak of a system of rules. Moreover, I also posit that rudimentary rules cannot be unequivocally individuated. In other words, one cannot provide a set of absolute criteria for rudimentary rule-individuation. Speaking metaphorically, the same 'amalgam' of rudimentary rules may be divided in various ways, giving different sets of fully individuated rules. What follows, at the rudimentary level there exists no absolute criterion for distinguishing kinds of rules. One is in no position to ascribe different rules to different categories (linguistic, legal, moral, mathematical). Rudimentary rules are relatively normatively unified.

The second view is that of *abstract rules*. They are abstract entities, which can exists in isolation (i.e., apart from a system of rules). What follows, is that such rules can be unequivocally individuated; one can also ascribe some rule-categories (legal, linguistic, etc.) to them. I also claim that the emergence of the abstract rules is the result of our reflecting on our behaviour. It is only through reflection that we can categorize different kinds of rules, ask what are the criteria for distinguishing legal, moral, linguistic, mathematical and other types of rules. Finally, in order to categorize abstract rules one needs to *express them in language*. Thus, abstract

rules are *linguistic entities*.

Now, the crucial step is to recognize that both stances are compatible. The ontology of rules proposed here is as follows. The regularities of social behaviour and some mental states *give rise* to the emergence of rudimentary rules. These, in turn, are the 'scaffolding' for the abstract rules.\(^\text{34}\) The proposed ontology is neither monistic nor dualistic. It is not monism, as rules cannot be reduced to facts, neither mental, nor social, nor a combination thereof. It is not dualism, as it is not assumed that there exists a separate and ontologically autonomous sphere of rules of behaviour. Moreover, the ontology of rules offered here fares well with the problems of the normative horizon. On the one hand, it is not confined to one particular kind of rules (moral, legal, mathematical, etc.). On the other, through the idea of rudimentary rules and their relative normative unity, it is safeguarded from the risk of leaving some kind of normativity unexplained.

-4-

There is an evolutionary scenario pertaining to the emergence of various kinds of rules that resonates well with the above-sketched ontology of rules. It was developed by Merlin Donald and Michael Tomasello, but is also argued for by Michael Arbib, Jordan Zlatev, and others. The fundamental observation motivating this stance is that the genetic difference between the human species and other animals is not so big (approx. 1 – 1,2%). This constitutes an argument for the thesis that the biological adaptation enabling the flourishing of human culture must be relatively 'small'. In other words, the proponents of the described scenario claim that it is impossible to account for the development of various aspects of culture (language, morality, science, etc.) by recourse to a large number of biological adaptations. In particular, Michael Tomasello believes that „the 6 million years that

---

\(^{34}\) Possibly, one can also distinguish between so-called *intermediary rules*, i.e. rules that exhibit the characteristic features of both rudimentary and abstract rules. These would be the 'rules we live by'. However, their introduction would result in much more complex conception – for the sake of readability I am not following this line of inquiry here. Cf. B. Brożek, „Normatywność prawa. Szkic teorii‖ (The Normativity of Law. A Sketch of a Theory), *Logos i Ethos* 1 (28) 2010, p. 29-66.
separates human beings from other great apes is a very short time evolutionary, with modern humans and chimpanzees sharing something on the order of 99 percent of their genetic material – the same degree of relatedness as that of other sister genera such as lions and tigers, horses and zebras, and rats and mice. Our problem is thus one of time. The fact is, there simply has not been enough time for normal processes of biological evolution involving genetic variation and natural selection to have created, one by one, each of the cognitive skills necessary for modern humans to invent and maintain complex tool-use industries and technologies, complex forms of symbolic communication and representation, and complex social organizations and institutions.”

Tomasello claims that “there is only one possible solution to this puzzle. That is, there is only one known biological mechanism that could bring about these kinds of changes in behaviour and cognition in so short a time (…). This biological mechanism is social or cultural transmission, which works on time scales many orders of magnitude faster than those of organic evolution.” This, in turn, is made possible by three forms of learning: imitative, instructed and collaborative, which are conditioned by „a single special form of cognition, namely, the ability of individual organisms to understand conspecifics as being like themselves who have intentional and mental lives like their own.”

There is a number of facts that underline the distinctive human capacity to relate to conspecifics. The strongest such line of arguments is connected to the differences between humans and other primates. In a number of experiments Tomasello and his colleagues have demonstrated that other great apes' learning differs substantially from human ways of cultural transmission. In particular, apes learn by emulation (i.e., they grasp only the means-ends structure of an activity and do not copy the pattern of behaviour), while humans learn by imitation and instruction.

These observations suggest that the ability to imitate is one of the crucial

36 Ibidem, p. 4.
37 Ibidem, p. 5.
adaptations in the evolutionary history of humankind. Tied with this is the capacity of 'mindreading' or 'intention-reading'. In addition, evolution has equipped us with a cluster of emotions (for Tomasello, guilt and shame are the basic emotions for cementing social bonds). These adaptations, taken together, are responsible for what Tomasello calls human mutualism.\(^{38}\) We not only have the ability to take the perspectives of others; we can also take a perspective with others. To put it differently, humans not only understand what some other individuals do (their intentions), but also do things together with them (i.e., we are capable of we-intentionality). This, according to Tomasello, is the key for understanding the possibility of cumulative cultural evolution. Moreover, mutualism explains, in an interesting way, human tendency to follow various types of social norms, and in particular the norms of cooperation (doing things for mutual benefit) and of conformity (doing things to identify with the group). On this account, our altruism is but a manifestation of a more general tendency to do things together (mutualism).\(^{39}\)

A similar evolutionary scenario was sketched by Merlin Donald in relation to the emergence of language. Donald claims that the sources of human ability to use language are based on mimetic skills, which evolved some 2 million years ago. He distinguishes between mimicry, imitation and mimesis. Mimicry is a simple copying of some action, with no understanding of its goal. Imitation is more abstract and flexible, as it takes into account the goal of the action. Finally, mimesis is defined as "the reduplication of an event for communicative purposes. Mimesis requires that the audience be taken into account. It also demands taking a third-person perspective on the actor's own behaviour. Some examples are children's fantasy play, the iconic gestures used in a social context, and the simulation of a 'heroic' death during a theatrical performance."\(^{40}\) Mimetic skills are thus founded on the ability to imitate, which in turn is conditioned by the mimicry skills.

Donald identifies four main types of mimetic representation, which are key to the transmission and propagation of culture: (1) reenactive mime, characteristic of

\(^{39}\) See ibidem, *passim*.
role-playing; (2) precise means-end imitation (as in learning how to fry an egg); (3) the systematic rehearsal and refinement of skill (as in learning how to drive a car); and (4) nonlinguistic gesture (as in learning how to dance). He claims further that these mimetic skills were the foundation for the emergence of language and all the other forms of culture. He stresses that his proposal differs from the traditional scenarios which condition the emergence of culture on the prior emergence of language (the language first theory). According to Donald some forms of culture, based on the mimetic skills, must have preceded language and enabled its evolution (the culture first theory).

Donald's theory leads to profound consequences. First, he claims that human mind is intimately linked to the society in which it flourishes. One can even say that it is co-created by the community. The communal practices are constitutive of the human mind, both in phylogenetic and ontogenetic dimensions. Second, language is not an individual but a network-level phenomenon: its evolution resembles the evolution of an ecosystem rather than of a single organism. Third, it follows that „cognitive neuroscientists are unlikely to find an innate language acquisition device, and should redirect their investigations toward the powerful analogue processing systems out of which language can emerge in group interactions.”

Such an evolutionary scenario supports the philosophical theory formulated above, according to which rudimentary rules supervene on both mental states and regularities of social behaviour, as well as the claim that rudimentary rules are relatively normatively unified. On Tomasello and Donald's view the ability to imitate leads to the emergence of a complex cluster of social regularities. Moreover, this ability is non-discriminating in relation to the type of behaviour (thus one cannot say that humans tend to learn moral behaviour easier or faster than other kinds of action). Finally, in its ontogenetic dimension, the imitation/mimesis hypothesis suggests that the acquisition of language does not consist in learning isolated words, but rather simple expressions (holophrases). If so, it is difficult to distinguish between spontaneous language-learning and the learning of other kinds of social rules; e.g., the same behaviour may count both as language- and morality-relevant.

41 Ibidem, p. 294.
In recent years, there have been much interest in the neuroscientific study of rule-guided behaviour with focus on the role of the prefrontal cortex (PFC). Various regions of PFC are implicated in a number of mental functions; in particular, it is believed that the lateral PFC regulates behaviour and controls response to environmental stimuli, while the orbitomedical frontal cortex, with its neurocortical and limbic connections, integrates information about emotions and memory related to environmental stimuli. The role of the PFC is prominent in action imitation, gestural communication, or monitoring of orofacial and hand actions (ventrolateral cortex), as well as in visual and auditory object identification and speech production.

It is speculated that the development of the PFC in children is strongly correlated with their ability to follow rules of different complexity. A number of psychological experiments have indicated that newborns are incapable of rule-following. Infants (approx. 9 months old) can repeat simple actions, but cannot adjust their behaviour to changing circumstances. Only at the age 2-3 children are capable to follow a simple conditional rule ('When the light flashes, press the button'). Further development enables children to work with two conditional rules at the same time, and only at the age of 5 they are capable of switching between two rules during a task.

Zelazo speculates that the development of the complex rule use in children mirrors the development of the levels of consciousness. Moreover, he believes that language plays a key role in rule use. First, the formulation of rules is hypothesized to occur primarily, if not exclusively, in potentially silent, self-directed speech. People need to talk their way through rule use tasks – and more generally, through problems requiring conscious control. (...) Second, the use of language, and in

particular, labelling one’s subjective experiences, helps to make those experiences an object of consideration at a higher level of consciousness (within developmental constraints on the highest level of consciousness that children are able to obtain). (...) This (...) is consistent with the hypothesis that labelling their initial subjective perspective places children at a higher level of consciousness, from which it is possible to reflect on their initial perspective, and from which it is easier to access an alternative perspective on the same situation."\textsuperscript{44}

Finally, Zelazo observes that "the function of PFC is (...) hierarchical in a way that corresponds roughly to the hierarchical complexity of rule use (...). As individuals engage in reflective processing, ascend through levels of consciousness, and formulate more complex rule systems, regions of lateral PFC are integrated into an increasingly elaborate hierarchy of PFC function via thalamocortical circuits."\textsuperscript{45} It is also consistent with the fact that the order of the acquisition of more complex rule types corresponds to the maturation of the relevant regions of PFC. The volume of grey matter reaches its adult level earliest in the PFC regions which are 'responsible for handling' simple conditional rules.\textsuperscript{46}

The conception proposed by Zelazo strongly indicates that rule-use develops in humans over time. He indicates that in order to acquire the ability to follow complex cluster of rules a child needs high level of consciousness as well as language skills. However, the focus on the role of PFC in rule-guided behaviour is only one part of the story. In particular, it helps little to understand the neural mechanisms for pattern recognition and re-enactment.

The single most important discovery that brings us closer to the understanding of how pattern recognition and pattern re-enactment are possible is undoubtedly the discovery of the mirror neurons by the neuroscientists from Parma in the early 1990s. Mirror neurons fire both when an action is executed and when it is observed. So far, the role of mirror systems has been implicated in, at least, action recognition, gestural communication, language, empathy and other emotional responses, and – in particular – in imitation and 'mindreading'.

\textsuperscript{44} Ibidem, p. 449. \\textsuperscript{45} Ibidem, p. 451. \\textsuperscript{46} Cf. ibidem.
Giacomo Rizolatti claims that there are two types of mirror neurons: high level resonance and, in consequence, two types of imitation. The high level resonance is used to mirror the goal of an action, while the low-level resonance copies the way of acting. According to Rizolatti only human brain takes advantage of both mechanisms, which enables imitation sensu stricto. In apes and (possibly) other animals only high-level resonance is used, and this explains apes' learning by emulation only. The utilization of both high-level and low-level resonance plays a key role in the enhancement of the flexibility and stability of human social reaction, as it enables – through recombination – to use the same patterns of behaviour, learned by imitation, to realize different goals, or to realize one goal with different means.47

An intriguing account of the role of mirror systems is proposed by Vittorio Gallese, who advances the shared manifold hypothesis. The 'manifold' constitutes a simple, intersubjective space of information, based on the mirror mechanism. The mode in which mirror neurons operate suggests that the brain encodes 'own' behaviour in the same way as the behaviour of others. This leads to the emergence of a primitive form of intersubjectivity:

The shared blended space enables the social bootstrapping of cognitive and affective development. Once the crucial bonds with the world of others are established, this space carries over to the adult conceptual faculty of socially mapping sameness and difference (“I am a different subject”). (...) The shared space provides an incredibly powerful tool for detecting and incorporating coherence, regularity, and predictability in the course of an individual’s interactions with his or her environment. (...) The concurrent development of language probably contributes to further separating out of single characters or modalities of experience from the original multimodal perceptual world, but the shared intersubjective space does not disappear. It progressively acquires a different role: to provide our self with the capacity simultaneously to entertain self–other identity and difference. (...) My proposal is that the "selfness" quality we readily attribute to others, the inner feeling of "being like me" triggered by our encounter with others, is the result of this preserved blended intersubjective space. Self–other physical and epistemic interactions are shaped and conditioned by the same body and environmental constraints.48

---

48 V. Galese, "<<Being Like Me>>: Self-Other Identity, Mirror Neurons, and Empathy", in S. Hurley, N.
Thus, Gallese claims that both phylogenetically and ontogenetically the shared manifold is prior to the 'self-other' distinction. This proposal, even if only roughly correct, indicates that a primitive, biologically conditioned form of intersubjectivity emerges early in phylogeny and ontogeny. In consequence, one may argue for the thesis that both in individual development, and in the evolution of our species, the social dimension in perception is prior to the emergence of language, culture and even a fully functioning mind.

Against such a framework, it is interesting what is the role of mirror neurons in the development of human mind. The key ingredient here is the ability of 'mindreading' or attributing intentions to others. There are two competing conceptions of the emergence of 'mindreading' in relation to imitation: the theory theory and the simulation theory. The former is forcefully defended by the developmental psychologist, Andrew N. Meltzoff. He speaks of three stages in the development of the ability to understand others. The first stage – in which imitation plays an eminent role – consists in the infants' ability to relate their behaviour to the similar behaviour of other people. At the second stage some of the own actions are tied up with some of one's own inner experiences. Finally, the third stage consists in connecting the behaviour of others with their inner experiences. The connection is based on an analogy with own actions and experiences. Thus, Meltzoff suggests that the attribution of mental states, intentions, etc. is possible through a kind of reasoning from the 'first-person' to the 'third-person' relations. To put it differently: the attribution of mental states to others is conditioned by the development of a 'what-does-it-mean-to-have-inner-experiences-theory', which is constructed on the basis of the 'first-person' perspective.49

A different approach is taken by the proponents of the simulation theory. For example, Alvin Goldman believes that the attribution of mental states to others may be depicted as follows. At the first stage, through the mirror systems, our mind generates a state similar to the mental state of the observed person. At the second stage this information is processed by one's own neural architecture working 'off-

---

line'. In other words, Goldman claims that the brain uses the same circuits when we experience some mental states as well as when we observe someone else experiencing them. The result of such 'off-line' simulation is attributed, at the third stage, to the other person. It is worth stressing that on this account the 'reading' of the mental states of others requires no 'inferences'.

The above presented conceptions are connected to another controversy: what is phylo- and ontogenetically prior: 'mindreading' or the ability to imitate? The priority of 'mindreading' is defended by Rizolatti, Tomasello and Carpenter. The opposite stance is taken by Meltzoff, Goldman and Gordon. According to Susan Hurley and Nick Chater those two views are not incompatible. They believe that 'mindreading' and imitation develop simultaneously: imitation *sensu stricto* requires complex 'mindreading', and *vice versa*. However, such 'full blooded' abilities are developed on the basis of simpler skills:

Very early imitation may express a fundamental self–other similarity, while the distinctive human capacity for imitative learning with its flexible means-ends structure in turn contributes to the development of the self–other distinction and of more advanced mind-reading skills.50

My final comment concerning the neural architecture of rule-following pertains to the growing body of theoretical work on the 'automatic' or 'quasi-automatic' character of much of human behaviour. It is argued that a majority of human actions are regulated and executed at the unconscious level. In this process, the key role is played by our emotions. However, it is also stressed that much of our regular behaviour is *inculturated*. This resonates well with the description of the functions of mirror neurons presented above. For example: in an often quoted article, „The Emotional Dog And Its Rational Tail“, Jonathan Haidt speaks of the intuitive character of the vast majority of human morally relevant decisions. He believes that our 'moral intuition' is not innate: it is rather shaped by social interactions and consists in re-enacting learnt patterns of behaviour. The conscious or reflective use of moral rules is rare – reasoning serves rather for an *ex post factum* rationalization.

of the decisions made unconsciously. A similar story may be told in relation to 'embodied' language and communication rules, tool-use, etc.

From this, the following picture of 'rule-following' emerges. Firstly, one can speak of rule-guided behaviour. In such relatively rare cases the decision as to how to act is preceded by conscious reasoning which aims at reconstructing possibly applicable rules of behaviour, their comparison, weighing of reasons, etc. Secondly, there exists rule-complying behaviour, which is not, however, motivated by rules. It is a large class of 'automatic' or 'intuitive' actions; in other words, these are 'embodied' actions, which replicate social patterns of behaviour. Although 'full-blooded' rules do not influence such behaviour directly, they are nevertheless implicitly 'present' through the process of social training. Moreover, the existence of the 'intuitive' behaviour seems a precondition for the development of abstract rules. The fact that through our ability to imitate we behave in similar ways and replicate unconsciously complex patterns of action constitutes a 'scaffolding' for the most complicated cultural artefacts.

I believe that the above sketched neural mechanisms of rule-following reinforce both my philosophical conception pertaining to the ontology of rules as well as the evolutionary scenarios of Tomasello and Donald. Recall my claim that rules are entities which depend (supervene) on both mental states and social interactions. The mirror mechanism hypothesis and the ability to imitate connected to it are fully consistent with the insistence on the social dimension of rule-following. Moreover, some features of the mirror system (well captured by the Gallese's notion of the shared manifold) imply a fundamentally social or intersubjective character of human cognition and behaviour. I posit also that it should be possible to distinguish between rudimentary and abstract rules. Rudimentary rules are simple, pre-linguistic and relatively normatively unified. This seems to find confirmation in the conception of an embodied rule of behaviour, which is non-linguistic, non-complex, and – given the extent of shared neural mechanisms behind the 'embodied' behaviour – relatively normatively unified. On the other hand, abstract rules are

linguistic, complex and normatively diversified. Again, it is consistent with the notion of a rule that we express in language and use for guiding our behaviour (i.e., rules we use consciously). At this level, one can reflect over one's behaviour, develop theories of how to act (normative ethics, law, theory of rationality) and, in consequence, classify various abstract rules as legal, moral, prudential, linguistic, etc.

-6-

Naturally, there is a number of well-established biological theories that are inconsistent with the above described conception of rules. There is no place in this essay to review all such proposals. Instead, I would like to concentrate on one particular cluster of ideas known as the 'modular paradigm' in cognitive science. If the mind is modular, and if among its modules there are, as often indicated, a 'language module', a 'normativity module' or a 'social contract algorithm', and if the modules are innate, than the thesis that rules have both mental and social dimensions is questionable. Moreover, the modularity view of the mind undermines also the normative unity thesis pertaining to the rudimentary rules. I believe, however, that the modular paradigm is either false or trivial in the sense explicated below. Of course, even if I succeed in demonstrating the futility of the modular approach, it is still not enough to vindicate the evolutionary/neuroscientific theory I defend. However, such a rejection of modularity would increase the plausibility of my proposal.

The thesis pertaining to mind's modularity was forcefully argued for by Jerry Fodor in 1983.\(^{52}\) He claimed that mind consists of independent modules which are:

- localized (i.e. their functions are realized in a dedicated neural architecture);
- subject to characteristic breakdowns;
- mandatory (i.e., they operate in an automatic way);
- fast (i.e. they generate outputs quickly);

---

- shallow (i.e. they have simple outputs);
- ontogenetically determined (i.e., they develop in a characteristic pace and sequence);
- domain specific (i.e., they process a restricted class of inputs);
- inaccessible (i.e., higher levels of processing have limited access to the representations within a module);
- informationally encapsulated (i.e., modules cannot be guided by information at higher levels of processing).

The above mentioned features characterize what may be called 'strong modularity thesis'. There is little doubt that on the strong reading the modularity thesis is plainly false. The mind is not a mechanism consisting of modules which display simultaneously the above listed features. Moreover, there is evidence that there are no (or there are but a few) mental mechanisms that may be ascribed some of those properties. This claim is vividly defended by Jerry J. Prinz,53 who observes, first, that the neuroscientific data does not support the thesis pertaining to the 'strong' localization of mental functions in the brain. Although there is some level of localization, one cannot maintain that any important functions are realized by dedicated neural architecture. Second, Prinz indicates that the properties of automaticity, quickness and shallowness cannot serve the identification of an interesting class of mental operations: e.g., not all automatic processes are fast or shallow (consider circadian rhythms regulation or semantic priming). Third, Prinz questions the ontogenetic determination thesis and the associated claim that 'modules' are innate: the variations in the development of various mental functions in children – both cross-cultural and individual – strongly indicate that there are no innate modules responsible for language skills or 'mindreading' or understanding 'folk physics'. Fourth, Prinz raises serious doubts regarding domain specificity of modules. On the one hand, he argues that both the notion of domain and of specificity are quite vague and arbitrary; on the other, he points out that the 'best candidates' for domain specific modules – such as language module or

‘mindreading’ module – share vast and important fragments of neural infrastructure with other mental faculties. Fifth, Prinz criticizes Fodor’s arguments for informational inaccessibility and encapsulation of the modules. Fodor’s case was based on a certain interpretation of perceptual illusions such as Müller-Lyry test. Prinz claims that these phenomena may easily be interpreted in a different way; he also mentions a number of top-down effects, i.e. situations in which our beliefs (and other higher mental functions) influence our perception.54

If Prinz’s arguments are sound (as I believe they are) Fodor’s conception of the modularity of the mind is plainly false. Moreover, the above arguments suffice, in my opinion, to reject some moderate refinements of Fodor’s position (e.g., the evolutionary psychologists’ conception of the modular mind). Of course, one can always claim that there exist some components of some neural systems that perform dedicated functions, are fast, shallow, automatic, ontogenetically determined, and so on. This, however, would be trivial – such a conception of modules would have no explanatory or heuristic value.

A nice illustration of the problems surrounding the modularity claim is the social contract algorithm hypothesis formulated by Cosmides and Tooby55. It is sometimes believed that the apparent existence of such an algorithm has far-reaching consequences for our understanding of what is morality and law. In order to assess the merits of Comides’ and Tooby's conception, it is necessary to begin with an exposition of the basic assumptions of evolutionary psychology. They are:

- both our biological and mental mechanisms are evolutionary adaptations generated by the natural selection in the ancestral environments;
- our mental mechanisms are domain-specific calculation modules. Thus, human mind is not a tabula rasa. It consists of mechanisms (modules) which are dedicated to solving specific adaptive problems.

Both these claims are controversial. The former says that our basic mental mechanisms were formed in the ancestral environments, i.e. in Pleistocene (2 million - 10 thousand years ago). This is followed by the claim that the mechanisms

54 Cf. ibidem.
in question have remained fairly unchanged and they will not change in the future. Such an account minimizes the influence of culture on human behaviour, although it does not exclude it altogether.⁵⁶

The social contract algorithm is such a mechanism that evolved in Pleistocene and is independent of cultural influences. The key argument justifying the existence of the social contract algorithm is the *Wason selection task*. It is an experiment that aims at demonstrating that there are huge differences in performing normative and descriptive tasks by humans, even though both kinds of tasks have the same logical structure⁵⁷. In a simple test which uses the *modus ponens* and *modus tollens* rules of inference the participants were asked to make inferences on the basis of the following conditional: ‘If a person P has the Ebbinghaus Disease, she has troubles with memory’. There were only 26% of correct answers. When a normative conditional was used, utilizing exactly the same logical structure, ‘If you borrow my car, you will need to refuel it’, the level of correct answers reached 76%. This is interpreted as indicating that in solving both tasks the participants used different mental mechanisms; the social contract algorithm, applied in the normative version of the test, is claimed to be faster and more accurate than the slow and fallible general abstract thinking abilities.

This conclusion is groundless. As shown in the experiments of Sperber ad Girotto, one can frame such descriptive tasks in which there are more than 50% of correct answers. More importantly, they were able to construct normative tests in which the percentage of incorrect answers exceeded 80%⁵⁸. Their diagnosis is that the percentage of the correct answer in the *Wason Selection Task* does not depend on whether the conditional in question is descriptive or normative; rather, decisive is the way of framing the problem. More precisely, they claim that the more relevant for the asked question is the conditional used in the experiment, the bigger is the number of correct answers. In other words, Sperber an Girotto believe that the

---

⁵⁶ Cf. ibidem, p. 52.
selection task used by Cosmides and Tooby has nothing to do with reasoning; it is rather a test of understanding. A relatively small number of correct answers in the descriptive version of the test results from the fact that the participants do not understand it, when in the case of the normative one the level of understanding is much higher.

In addition, the modularity thesis is subject to a number of philosophical objections. Let us assume, for the sake of argument, that there really exists a 'normative module' or 'the social contract algorithm'. How does it help us to elucidate the concept of law? How does it contribute to answering the ontological, normativity and psychological questions? It depends on the way we understand the operations of the alleged module. The strongest, but controversial (to say the least) interpretation is given by Robert Nozick. In Invariances he says:

The capacity for following norms and the predisposition to follow some norms or other – norms of the local group, of young peers, of our parents and elders – is innate. It will be helpful to consider this as a specialized capacity, a 'normativity module'.

This view is reinforced in the following passage:

So let us suppose that there has been evolutionary selection for our normative capacity and receptiveness. There has been selection for a normativity operator within our cognitive and emotional apparatus, an operator that attaches an internal 'ought' to certain behaviour or patterns of behaviour. Some things ought to be done, other things ought not to be done. We learn these rule-governed patterns and we behave in accordance with them, not solely or mainly because of incentives, because of carrots and sticks, but because we are not by nature unruly creatures. We are, in fact, ruly ones.

If there were any doubts, as to whether Nozick takes the 'normativity module' to explain the psychological dimension of rule-following, or rather he defends a stronger thesis that the module is the source of normativity, the following quote seems to give a definite answer:

An ought cannot be derived from an is; normativity cannot be derived from descriptiveness. Yet the descriptive fact that we do have biologically instilled normativity boxes and operators (…) can be given a thoroughly naturalistic and non-normative explanation.

Thus, Nozick seems to believe that it is the normative module, one shaped by the

60 Ibidem, p. 271.
evolutionary processes, that solves the normativity puzzle. This is a very strange conception, one that replaces a puzzle with a mystery. Nozick's 'normativity boxes' are only a new name for an old problem, and to say that they are a product of evolution is of little value – it is close to a tautology.

A more modest reply to the question what is the role of the 'normative module' vis a vis the problem of rule-following is to say that its existence constitutes the answer to the psychological question. The fact that we are equipped with dedicated architecture for social reasoning, which is an important ingredient of our ability to cooperate, provides an explanation for our tendency to follow rules. Observe, however, that this says nothing in reply to the normativity question. Therefore, a proponent of the moderate interpretation of the role of the 'normative module' has two options available: she can either reject the problem of normativity and opt for reductionism, or seek the answer in a sphere which is impenetrable by the methods of science. In the former case, the problems surrounding Petrażycki's solution re-emerge (i.e., there is a need to postulate the existence of at least one rule that has the 'normative force'). In the latter, one needs to follow Kant's metaphysics in saying that the ultimate sources of normativity lie in the noumenal sphere, one that is beyond the reach of science.\textsuperscript{62} I believe that – as Wittgenstein suggested – these conclusion may be avoided once one takes the social dimension as a constitutive element of rule-following. This, however, is difficult to reconcile with the modularity paradigm.

\textsuperscript{-7-}

It is not my claim that the 'biological' picture of rule-following sketched above is true or even \textit{roughly} true. However, I believe it to be quite interesting that between the three accounts of rules I offered – philosophical, evolutionary and neuroscientific – there is a fair degree of coherence. I think it is a powerful

\textsuperscript{62} Such a position seems to be defended by Wojciech Załuski in his \textit{Evolutionary Theory and Legal Philosophy}, Edward Elgar: Cheltenham – Northampton 2009, pp. 102-105.
argument in favour of the cluster of hypotheses I presented.

Let me briefly pause here and summarize my answers to the three questions posed at the beginning of this paper. As regards the ontological question, I believe that rules (not only legal, but any kind of rules) *supervene* on mental states and social interaction. More precisely, rudimentary rules result from mainly unconscious operations of the mind and the social transmission of the patterns of behaviour, while the abstract rules – whose very existence is possible only *vis a vis* the system of rudimentary rules – are expressible in language and, as such, supervene on our conscious acts of reasoning and the social development of normative systems (formulated in language). The rudimentary rules have a motivational force through our tendency to cooperate and conform to the group. The mechanism of being motivated by abstract rules is more complicated. On the one hand, it is possible that the process of reflecting upon an abstract rule leads to choosing a given course of action; on the other – some abstract rules, when commonly followed, are propagated through the pattern recognition/pattern re-enactment mechanisms (i.e., they become rudimentary rules or clusters of rudimentary rules). Finally, rudimentary rules have a primitive normative force: they are objective ('shared') patterns of behaviour, and although they are followed 'blindly' or 'automatically', they are subject to social appraisal and correction. The abstract rules, on the other hand, may be said to constitute objective reasons for action as they may be subject to discussion; they may also be invoked as reasons for some particular course of action. However, whether an abstract rule is considered an objective reason for action or not depends on the theory of rationality one entertains. In other words, the notion of 'the objective reason for action' is not absolute: it is relative to the accepted standard of rationality (e.g., the same rule may be considered a reason for action from the perspective of the rational choice theory, while being no such reason according to the Kantian conception of practical rationality).

Now, if this picture is (partly) right, it provides an original answer to the question 'What is law?' It is clear that law cannot be defined at the level of rudimentary rules, i.e. the 'embodied' and normatively unified kind of rules. Anyone looking for a working definition of law must operate at the level of abstract rules. At
this level, however, there seems to be no definite criterion as to what counts as law. In other words, there is no single such definition. Law may be defined in various ways, depending on our cognitive or practical needs. The Holy Grail of legal philosophy, i.e. the 'true nature of law', is apparently nowhere to be found.

There is one more corollary of my conception I would like to mention. Most of our behaviour is rule-observing: we do much things 'automatically' or 'unconsciously', but these actions may count as complying with the precepts of the law. For example: when I buy a bus ticket or cross the road at the green light I usually do not reflect upon a relevant legal rule – I do what is 'embodied'. This counts, however, as acting in accordance with the existing law. Of course, it is also possible that I use legal rules ('legal' according to some definition of law) as guidance. When I file a tax return or when I prepare some document for the court or sign a mortgage agreement, I may contemplate and reflect upon legal rules, and this reasoning may motivate or influence my decision as to the applicable course of action.

Finally, the above discussion may serve as the basis for formulating a number of conclusions pertaining to the 'philosophy in biology' method. When 'philosophy meets biology', i.e. when philosophers are willing to use the findings of evolutionary theory or neuroscience to analyse philosophical problems or develop philosophical theories, the following caveats must be kept in mind. First, the structure of biological theories differs substantially from that of physics. Of course, one may defend a view that at the most general level biological theories consist in formulating a hypothesis which then is tested in experiments, and, in light of their results, rejected, modified or theoretically developed and subjected to further experiments. However, this picture is very general. At a more concrete level, the characteristic feature of biological theories (at least evolutionary theory and neuroscience) is their multidisciplinarity. For example: in Tomasello's conception the use is made of the findings of developmental psychology, primatology, evolutionary theory, neuroscience, comparative anatomy and evolutionary game theory. In the process of theory-construction three criteria are used: empirical adequacy, convergence and coherence. The sole criterion of empirical adequacy would be
useless, as there are usually many ways of interpreting the given set of experimental data. The criterion of convergence is applied in situations when the experimental data from various sources simultaneously support the same theoretical hypothesis. A nice example is the claim that the human ability to imitate exceeds the corresponding skills in great apes. This hypothesis is backed by the results of psychological experiments (with the participation of children and apes), as well as the comparative analysis of the structure of human and apes' brains. Coherence, in turn, is a relation between two (or more) theories: the more coherent they are, the higher is the degree in which they fulfil three logical conditions: of consistency, meaningful inferential connections and relative unification. For example: Donald's evolutionary scenario is highly coherent with the theories pertaining to the role of mirror neurons in human mimetic skills.

Second, biology often makes use of interpretation paradigms. One such example is evolutionary psychology, which interprets the results of experiments as if the human mind consisted of biologically determined modules which evolved in Pleistocene. Third, it follows that the borderline between biological (evolutionary, neuroscientific) theories and philosophical conceptions is 'blurred': sometimes it is difficult to say whether the given theory is still 'biological' or already 'philosophical' (consider, e.g., Zelazo's conception of the development of the levels of consciousness or even Gallese's 'shared manifold hypothesis').

On the other hand, when 'biology meets philosophy', the former may take advantage of the skills of the latter. In particular, philosophy may serve to clarify the mechanisms and criteria involved in the construction of biological theories. It may also uncover the hidden 'philosophical' presuppositions of biological theories or provide some help as regards the measure of convergence and coherence thereof. On some occasions philosophical conceptions may also serve as additional 'selection mechanisms' for choosing from among competing interpretations of experimental data (as, I believe, is the case with Wittgenstein's remarks concerning rule-following).