Popper and the Rationality Principle

by
Maurice Lagueux
Université de Montréal

Though Karl Popper's short paper on the rationality principle may not be the most frequently discussed of all of his writings on epistemological matters, it is very probably the most radically criticized. The fact that this champion of falsifiability suggested in this text not to reject a principle that he emphatically declares false has always been a source of embarrassment for his disciples and has often been characterised by his adversaries as a rather shameful theoretical development. In the present paper, I would like to show that, in spite of this fact, Popper's views on rationality, while at moments somewhat awkwardly formulated, are much more sensible than it is usually acknowledged and that they might even be considered as one of his most interesting contributions, and surely as his most underestimated one.

The nature of the rationality principle

But let us first recall what these views are. While frequently referring to reason and to rationality, Popper devoted only a few pages to the analysis of the rationality principle as such. It is well known that his ideas on the matter reached the philosophical community through the 1967 publication in French of a short paper entitled "La rationalité et le statut du principe de rationalité", the English version of which, unpublished until the 1980s, is now included in David Miller’s anthology entitled Popper Selections. Despite multiple criticisms of the views expressed in this paper, Popper never repudiated nor revisited it. However, in 1994, the year of his death, the text from which this paper on the rationality principle was derived was finally
published by M. A. Notturno in *The Myth of the Framework* under the title "Models, instruments and truth: the status of the rationality principle in the social sciences". A footnote at the beginning of this text confirms the fact that the paper published in French was indeed an extract of this original text, the latter being the result of a lecture delivered at Harvard in 1963. Properly speaking, the French paper on the rationality principle was not an extract of the original; it was rather a slightly extended rephrasing of only those ideas concerning the principle of rationality found in the earlier paper.

Nonetheless, given that the earlier text presents some points slightly differently, it may help to clarify the later text on the rationality principle. A first point, one which generated considerable debate, concerns the very nature of the rationality principle. Popper proposed his own version of this principle. His basic formulation is the following: "agents always act in a manner appropriate to the situation in which they find themselves" (Popper 1985, 361). This formulation was roughly the same in the initial lecture, according to which the principle requires that "the various persons or agents involved act *adequately*, or *appropriately* — that is to say, in accordance with the situation". (Popper 1994, 169) With this version of the rationality principle in mind, Popper has no trouble in claiming that such a principle is clearly false and that, consequently, it is anything but a priori. Who would deny that people frequently act in ways that are not appropriate to the situation? Popper proposes his famous example of the "flustered driver" who, by trying stupidly to park his car in evidently too small a space, *manifestly does not act in a manner which is appropriate to the situation in which he finds himself*. Many would even observe that such inadequate behaviour associated with neurosis and other forms of abnormal behaviour is far from being rare among human beings.
Nonetheless, Popper maintains that the rationality principle is still a "good approximation" of what takes place in human behaviour. He argues that cases of neurosis have indeed been explained by Freud and other psychologists, precisely with the help of their own version of the rationality principle. In such cases, however, the principle was presented in a significantly different version since typical responses of neurotic people are described as appropriate to the situation "as they see it." (Popper 1985, 363). The important modification introduced with this new version of the principle was underscored by critiques (e.g. Nadeau1993), who concluded that a principle thus construed can no longer be false and is clearly a priori, in a manner similar to Ludwig Von Mises' version of rationality principle. It is well known that the latter is a priori since it can allow one to present as rational any action whatsoever, since actions can always be described as appropriate responses to situations seen in one manner or the other (on this, see Lagueux 1995). Was Popper careless enough to refer to an a priori version of a principle which, only a few pages earlier, he emphatically claimed to be in no way a priori?

In a footnote to one of the paragraphs added to the text of his earlier conference in order to answer certain objections, Popper acknowledges without the least hesitation that he refers successively to two versions of the principle and he even submits a third intermediate version according to which the situation is said to be "as the agent could (within the objective situation) have seen it". (Popper 1994, 183) However, even if he were fully aware of the fact that the implications of the principle differ significantly according to the version adopted, Popper formally denies that any of his three versions make the principle a priori. He rather claims that "we sometimes act in a manner not adequate to the situation in any of the senses (1), (2) or (3)". (Popper 1994, 184) Strictly speaking then, whatever the version in which it is formulated, the
principle is false and in none of them is it a priori. According to Popper, even versions (2) and (3) are false since they are also falsified by the case of the flustered driver. The latter sees perfectly well that he cannot park his car in such a small space, but he is irritated to such a point that he makes desperate manoeuvres in order to park it there anyway, manoeuvres which, afterwards, require him to struggle to drive the car out of the cramped space into which it was needlessly squeezed. The driver behaves in a manner which is inappropriate to the situation even as he himself sees it. In contrast, the pedestrian who throws himself in the way of an oncoming cyclist in order to avoid being hit by a car acts in a way which is appropriate to the situation as he himself sees it. There is no doubt that if the pedestrian had a better view, he could have also seen the bicycle and thus avoid both accidents; however, taking into consideration what he actually sees, according to version (2), his response to the situation is completely appropriate. But since it cannot be excluded that, in similar circumstances, another pedestrian who was as aggravated as the flustered driver could, without any understandable reasons, choose to remain in the way of the car, the principle is false and not a priori. Thus, the claim that in almost all cases people behave in a rational way (understood in this sense) is based on an empirical knowledge. Consequently, the rationality principle according to which people always behave in such a way is false (and, thus, not a priori) while approximately true.

But can we really say that the flustered driver acts as he does without reason? Are we certain of the fact that he does not see the situation in a way which renders his behaviour relatively appropriate? In a previous paper (Lagueux 1993), I argued that he may, for example, derive satisfaction from the fact that his conduct demonstrates to the surrounding population that a city in which it is impossible to park one’s car is a city badly administered. On this basis,
I concluded that the main shortcoming of Popper's thesis was that its psychological account of what might be going on in the mind of a flustered driver was insufficient and that this example, after all, was probably not the most convincing illustration of irrational behaviour. However, the search for a better example should unavoidably lead one to the conclusion that no such example can do the job, precisely because an arbitrarily imagined example can always be arbitrarily construed in such a way that it associates the behaviour involved with a relatively appropriate reason. Even psychologists who have experimentally challenged the rationality of human behaviour tend to invoke another form of rationality to explain their findings. For example, in a famous paper documenting the fact that people’s decisions are frequently influenced illogically by the ‘frame’ in which the situation is presented, Amos Tversky and Daniel Kahneman explain that people are not ‘necessarily irrational’, since the incriminating behaviour can at least sometimes ‘be justified by reference to the mental effort required to explore alternative frames and avoid potential inconsistencies’ (Tversky and Kahneman 1986: 138). This is consonant with Popper’s view according to which testing a theory is testing a theory as a whole, in such a way that it is always possible to conclude that any alleged refutation of rationality instead refutes only a theory consisting of a model plus the rationality principle. If this is true, modifying the model (for example, including in it the compulsion to denounce the city administration) may be sufficient to save the rationality principle. Clearly, for Popper, the rationality principle is false not by virtue of having been formally falsified, but because it would be extravagant to attribute rationality to any action of any agent unless doing this on an a priori basis, which is judged groundless. Therefore, the point is not to find a better example of irrationality but to decide at a more general level whether one is correct in claiming a priori, as Ludwig von Mises did, that all actions are rational simply because they are purposeful.
Naturally, if one considers that an action is purposeful (and therefore rational) by definition, such a claim can be established a priori, but the resulting tautology can hardly be the basis of an empirical science.

For Popper, social sciences must be empirical sciences, but they must nonetheless be based on a rationality principle that stipulates that, in fact, human beings are constituted in such a way that they usually adapt their actions to what they see as their own interest. Considered in this fashion, this principle cannot be characterised as a priori. Let us try to illustrate why with the help of a thought experiment. In order to eliminate any confusion between a priori reasoning and introspection (or internal experience), let us consider an inhabitant of Sirius who is capable of deduction and induction but unable to experience self-interest. Arriving on planet Earth, this newcomer would be unable to conclude through a priori reasoning that (1) human beings are constituted in such a way that, depending on the way they see their situation, they can evaluate what is called their self-interest and that (2) they systematically tend to adapt their actions to what is required by that self-interest. At first glance, such newcomers from outer worlds might sooner think that human beings are absolutely irrational (or rather non-rational) organisms in a way similar to which most plants and animals appear to us. However, through careful observation of the regularities in human behaviour, they could arrive at the empirical conclusion that this behaviour is goal-oriented and not mechanically commanded, and from this they could induce that human beings are rational. With still more careful observation, they could even observe that very often human beings see their situation in an erratic way and that, in these circumstances, their (still rational) behaviour is adapted accordingly. Occasionally, however, they would have no other choice but to acknowledge that human behaviour is such that it appears irremediably stupid. Given these conclusions, they could, if their aim were to
understand human behaviour, adopt empirically a rationality principle even if experience has convinced them of the fact that this principle is approximately but not universally true. Such is the kind of principle that, according to Popper, is required by any model devised by social scientists.

If we grant Popper the empirical character of each version of his principle, we may raise the following question: how can Popper modify so freely the very meaning of a principle deemed to be so fundamental, in such a way that he can use simultaneously three versions of it? The answer must be looked for in his notion of model. According to Popper, in its very fabric, the model of an explanatory theory includes, in typified form, all initial conditions characterising the situation in which the phenomenon to be explained takes place. This means that what the agent sees may or may not be considered as part of the objective situation that the model describes. Therefore, from a formal point of view, all three formulations of the principle are not as different as they appear, since they differ only in their adaptation to the variable extension of the knowledge that is included in the relevant model describing the situation. The first version supposes that the agent has true knowledge; what Popper calls the second version implies that the alleged knowledge is partially wrong; and the third version corresponds to an intermediate case. In all cases, the rationality principle claims that the agent will act in a way appropriate to the state of this knowledge. For Popper, the important point is that, regardless of whether or not this model includes the (possibly incorrect) way in which the agent sees the situation, the model cannot work by itself. It needs to be animated by something else, in the same way that a mechanical model explaining the movement of the planets has to be animated by "Newton's universal laws of motion" (Popper 1985, 358). But, with almost all models
implemented in social sciences, the animation is provided by the rationality principle. Thus, a model which includes only objective aspects of the situation needs to be animated by the basic version of the principle according to which "agents always act in a manner appropriate to the situation in which they find themselves". But if one also includes subjective aspects of the situation in one's model, more precisely if one includes in it a description of the way in which the agent sees the situation, then such a model needs to be animated by the version of the same principle according to which "agents always act in a manner appropriate to the situation as they see it". Since Popper is not obliged to decide on what extent the content of the model should be pursued, he is not obliged to decide on which version of the rationality principle should be used, or, if one prefers, he is not obliged to decide to what kind of description of the situation this principle should correspond.

**Why immunise the rationality principle?**

For Popper, the important point is that an explanatory model in social sciences cannot work without the help of the rationality principle whose content, while empirical, tends to be minimal to the point of being "almost empty". Since refinements in social sciences model making imply that models include more and more detailed pictures of situations, the rationality principle tends to be reduced to the simple idea that the agent actually agrees with what is clearly presented by the model as the appropriate thing to do. It is clear that for Popper the rationality principle as such needs not include more than this, since all other physical and psychological aspects of the situation are preferably taken over in typified form by the model itself. In this context, it is not surprising that in both versions of his text, Popper claims that this minimal rationality principle "has nothing to do with the assumption that men are rational in
this sense — that they always adopt a rational attitude." (Popper 1985, 365) Since Popper takes
the trouble to specify "rational in this sense", it is legitimate to ask more precisely what sense
is in question. Clearly, given that Popper associates "this sense" with "a rational attitude", the
sense in question is one which he attributes to the notion of "rationality as a personal attitude"
in the paragraph preceding this last quotation: "Rationality as a personal attitude is the attitude
of readiness to correct one's beliefs. In its intellectually most highly developed form, it is the
readiness to discuss one's beliefs critically, and to correct them in the light of critical discussions
with other people." (Popper 1985, 365) a view which is closely connected to his "critical
rationalism". Thus, it is clear — and Popper does not seem to say anything else — that the
rationality principle (stating that individuals act in a manner appropriate to their situation) "has
nothing to do" with this tendency to correct oneself by criticism which, for Popper, constitutes
authentic rationality. Popper immediately emphasizes that, far from being as rich as this, the
principle in question is only a "minimum principle": "it assumes no more than the adequacy of
our actions to our problem situations as we see them" (Popper 1985, 365) or, to put it in the
terms that I used above, it assumes that the agent will agree with what is clearly presented by
the model itself as the appropriate thing to do.

But, as we have seen, nothing can assure us that the agent will actually agree accordingly,
even in such a situation; consequently, it is clearly false to maintain that agents will always
choose to act in the appropriate fashion. At this point, the question to be raised is one which has
so often been formulated as a decisive objection to Popper's view (for example, Hands, 1985):
why does the champion of falsificationism refuse to reject a principle that he himself says is
false? Given that Popper attributes to the rationality principle such a determinant role (from a
methodological point of view), one could be tempted to think that Popper's decision to immunise this principle against falsification is due to the fact that such a principle is nothing but a *methodological* principle somewhat similar to Popper's famous methodological principle, so convincingly defended in *The Logic of Scientific Discovery*, which stipulates that, given equally successful hypotheses, a preference must be granted to those which are the most easily falsifiable. In the two cases, we would be dealing with a methodological postulate whose empirical testing would be absurd. But such an easy way out is not accepted by Popper who rejects it in the most unequivocal terms. When called to say whether the rationality principle is a "methodological principle" or an "empirical conjecture", his answer is that "this second case is precisely the one that corresponds to my own view of the status of the rationality principle: I regard the principle of adequacy of action (that is the rationality principle) as an integral part of every, or nearly every, testable social theory"? (Popper, 1985, 361; see also Popper 1994, 177). He firmly maintains this point because he sees the rationality principle as an integral part of any empirical theory in social sciences, that is, as its animating part, just like the laws of motion are an integral part of Newton's astronomical theory. In contrast, a *methodological principle* cannot *be* a part of a scientific theory whose parts must be, according to him, empirical and hypothetical rather than a priori.

But if the rationality principle cannot be *itself* a methodological principle (or a methodological rule), the *decision* to immunise it can nonetheless be considered as based on a methodological principle. As Popper says, “if a theory is tested, and found faulty, then we have always to decide which of its various constituent parts we shall make accountable for its failure” (Popper, 1985, 362). The methodological decision has simply to be disentangled from the
rationality principle itself since, in contrast with the latter, the former is not part of the empirical theory. Thus to remove any trace of contradiction from Popper's views on rationality principle, it remains only to show that the methodological decision according to which the rationality principle (while being an empirical hypothesis) must be immunised against falsification is compatible with the methodological decision according to which preference must be granted to hypotheses which are the most easily falsifiable. However, presenting the decision to immunise the rationality principle as a methodological rule may look paradoxical since, according to Popper, the supreme rule “ says that the other rules of scientific procedure must be designed in such a way that they do not protect any statement in science against falsification” (Popper 1959, 54). But the point is that, in Popper’s mind, the principle of rationality is not protected against falsification by the decision to maintain it since, according to him, it is already falsified. Far from protecting scientific statements against falsification, this decision, according to Popper, allows us to falsify the other statements of the model, the only ones that are meaningfully falsifiable.

There is little doubt that it was in order to show that his methodological decisions were inspired by a consistent approach that, in a paragraph of fifteen lines, Popper expounded four arguments (Popper, 1985, 362) which refer to the fact that the fundamental role of the rationality principle is to help us to learn more about the world. For Popper, falsifiability is nothing but a means to promote our understanding by making theories testable rather than arbitrary and *ad hoc*. But these arguments suggest that, for the sake of understanding, dismissing the rationality principle is not recommendable even if it is not universally true, since replacing it by a principle admitting irrationality would open the door to arbitrariness and adhocness and would make true
understanding impossible. Indeed, if irrationality could be invoked in order to explain some social phenomena, any kind of irrational behaviour could be arbitrarily appealed to in an *ad hoc* fashion in the case of each unusual social phenomenon to be explained. In fact, serious social scientists have understood this point, and it is for this reason that almost all alternative theories proposed by them invoke one form or another of the rationality principle and exclude facile appeals to irrationality. For Popper this fact clearly illustrates that it is useless to reject the rationality principle if the goal of this rejection is to improve our understanding with the help of a better theory, since alternative theories worthy of consideration equally make use of the same principle. Moreover, according to Popper, falsifiability is an attractive idea because it makes possible testability and, consequently, better knowledge and better understanding. However, since different component parts of a theory cannot be falsified through the same test and since the theory to be tested is nothing but a model composed of a number of various elements to which an animating principle is added, when the test of a theory fails, one must choose between declaring false either one of the model's elements or the rationality principle that animates it. But, as Popper asks, what do we gain by blaming the rationality principle as such? We already know that it is false, or, at the very least, we know that it is not universally true. To observe, on the basis of various tests, that in some (relatively isolated) cases, it does not work, teaches us nothing that we did not already know. Nor would it provide us with information which could help us to improve our theory. In contrast, testing certain of the component parts of the model itself can be very instructive and very helpful if we want to modify our theory in order to improve it. It is in this fashion that we can improve our knowledge and our understanding. Popper seems to conclude that, given this situation, we loose nothing by maintaining a false principle whose rejection, in contrast, would leave us at a dead end.
Consequently, if our goal is knowledge and understanding, maintaining a (false) principle that allows us to construct progressively more accurate explanatory models would appear to be a good decision.

The particularity of the rationality principle

These arguments were designed in order to justify the methodological decision not to treat the rationality principle in the same way as any other (declared false) part of a theory. However, even if they highlight some of the non negligible advantages of immunising the rationality principle, one might easily remain unconvinced by them since they do not reveal the particularity of the principle that justifies its treatment in such a way. There is little doubt that if Popper avoids to particularising too much this principle, it is because, according to him, it plays exactly the same animating role that animating principles, like Newton's universal laws of motion, play in physicists' models. Indeed, before introducing the comparison between these two kinds of models, Popper says that "it is important to realise the close similarity of explanations in the social sciences with the explanations [...] in the natural sciences" (Popper 1985, 358). However, notwithstanding this parallelism, it is clear that Popper would agree that it would not make sense to immunise in the same way Newton's laws of motion if ever they were declared to be false. More important, even if Newton's laws animate the planetary model referred to by Popper, they do not animate all or almost all of the models constructed by natural scientists, whereas, according to Popper himself, it would be difficult to conceive of a model in the social sciences which would not be animated by the rationality principle. Thus, one must admit that the rationality principle occupies a place in the social sciences without an exact equivalent in the natural sciences. Popper was too committed to the unity of science to give too
much significance to this singular character of the social sciences and to characterise the rationality principle accordingly, but his "situational analysis" theory draws sufficient attention on some specific features of these sciences. Even if, for these reasons, Popper does not seek to determine what accounts for the exceptional role played by the rationality principle, other theoreticians have understood that this principle occupies an exceptional place because it is a condition of intelligibility of any phenomenon that derives from human action. As was clearly seen, for example, by Austrian economists like Menger, Mises, Hayek and Lachmann, a social phenomenon is explained when it is reduced to the consequences (usually unwanted) of human actions that are intelligible. And human actions are intelligible — they can be understood — when they are deemed to be rational, when they are motivated by reasons, which is to say when they are appropriate responses to situations as seen by those who make them. Were an action declared irrational in the sense of not being motivated by any reason whatsoever, such an action would clearly be unintelligible. How, indeed, could one pretend to understand an action and at the same time to declare it irrational or to declare it a totally inappropriate response to the situation even as the agent sees it?

But maintaining this principle after acknowledging that it is not universally true is simply to claim, as Popper did, that, in spite of the fact that strictly irrational decisions occur, human actions are nonetheless normally understandable. It is to claim that human actions are normally intelligible, not because they can be subsumed under a universal law, but because they can be subsumed under the principle of rationality. Being subsumed under this principle means that they are intelligible because they are rational. Natural phenomena cannot be explained in such a fashion because they do not derive from intelligible human actions. While reluctant to admit
that the social sciences could, for this reason, differ radically from the natural sciences, Popper clearly understood that the rejection of such a principle (for being not universally true) would lead to the rejection of the very possibility of understanding social phenomena.

For sure, one might prefer another type of science, which would not invoke notions like rationality and understanding. Accordingly, it seems that Popper's reluctance to push far enough the analysis of the specific character of the rationality principle is due to his own desire to see both the natural and the social sciences put on the same footing. However, while it is not strictly impossible that a science of social phenomena based on neurophysiological findings coupled with natural selection might be developed in the future without leaving any role for the rationality principle, almost all explanatory theories of social sciences developed up to now (or, at the very least, up to Popper's time) are based on this principle. Popper, who clearly understood this, managed to underscore as much as possible the parallelism between the natural sciences and this type of social sciences. To do this, he had first to depsychologise the rationality principle. Social sciences are based on understanding human actions, but this view should not, in his mind, be associated with either a Diltheyian Verstehen or a Collingwoodian knowledge of actions from inside. The tendency to do what looks appropriate, which is precisely what the rationality principle highlights, does not imply a kind of psychological introspection. It is for this reason that Popper reduced the rationality principle to a minimal principle requiring nothing more than an almost mechanical response to a situation. What allows Popper to say that such a mechanical response should be an appropriate one is not a psychological analysis of what is going on in the mind of the agent, but rather the fact that the situation is modelled in such a way that the model includes all the features which should
normally bring the agent — of whom it is only required that he attend to his own interest — to react in the very way that the model itself presents as appropriate.

Thus, what is required of the rationality principle is simply the idea that the agent is not stupid enough to avoid responding in a way which, given the situation as he sees it, corresponds to his own interest. The rationality principle thus construed is really "almost empty" as Popper says; in fact, it excludes only sheer stupidity. It is true that sheer stupidity exists in human behaviour — that we are convinced of this in the case of the flustered driver, or on the basis of any other consideration, matters little — and it is for this reason that Popper says that the rationality principle is false; however, such stupidity is far from being prevalent among human beings, and it is for this reason that Popper characterises the rationality principle as a "good approximation" of the truth. Popper did not associate explicitly his minimal version of the rationality principle with the non-stupidity of most human beings, but most compelling explanations provided by economics, throughout its history, can be reduced to theories based on a rationality principle understood in this minimal fashion.

To take only one of my favourite examples, consider Turgot who, as early as 1766, argued that a ‘current price’ has to prevail in a market (and not simply should prevail, as earlier ‘just price’ theorists maintained): ‘[...] if one [of the wine sellers] is not willing to give more than four pints for a bushel, the Proprieter of the corn will not give him his corn, when he comes to learn that someone else will give him six or eight pints for the same bushel’. (Turgot 1898, 29) Turgot's argument was clearly based on the fact that people are rational in the sense that they are not stupid. Who would be stupid enough to give one bushel of wheat in exchange for only
four pints of wine when it is well known that other wine sellers would be happy to give six or eight pints for the same bushel? Construed in this way the rationality principle implicitly but clearly used by Turgot is as minimal as the one proposed by Popper. It is clear that such a principle is false since there are always people stupid enough to give something in exchange of four pints even when they know that they can get six or eight pints for the same thing (and this even when the model is conceived of in such a way that altruism is excluded), yet it remains an approximation good enough to be used by almost all explanatory theories of economics. It would be easy to illustrate — though this has been done in a different paper (see Lagueux 2004) — the fact that almost all the explanatory theories of classical, Marxian, and Austrian economists are clearly based on such a principle, and that neoclassical theories are also based on it, even if these latter are usually modelled in such a way that it is strict maximisation which is presented to non-stupid agents as the appropriate thing to do. In each of these cases, models must be animated by a minimal principle which says only that people are normally not stupid enough to refrain from doing what those models present as the appropriate thing to do. Since almost all economic theories and most of social sciences theories can be construed in this way, it would appear reasonable to conclude that Popper’s highly criticized ideas on the rationality principle, far from being a shameful part of his work, are rather the result of a shrewd though somewhat sketchily exposed analysis of the state of social sciences.

**Quoted works:**


