447 *Etica & Politica / Ethics & Politics*, XXVI, 2024, 2, pp. 447-464
ISSN: 1825-5167

REASONS, ACTIONS, EXPECTED UTILITY, MOTIVATIONS

PIERPAOLO MARRONE
Università di Trieste
Departmento di Studi Umanistici
marrone@units.it

ABSTRACT
The paper deals with the received view, shared by many social scientists, ofexpected utility theory as a descriptive and prescriptive vision both of right action and of right procedure of thought. It states that in the classical formulation of the theory there are scarce hints for such a monistic interpretation, which does not pay justice to the pluralistic reality of the reason for action.

KEYWORDS
Reason, action, motivation, expected utility.

1. The reasons that may lead you to perform an action may of course be the most varied, but in some sense (e.g. on the basis of the background morality you accept) what you should believe is that the results of the action you will perform will maximise the utility you expect as a reward for having chosen that particular course of action rather than another. What other reason, after all, should have led you to perform it? Note that such a description of the reasons for action can, quite consistently, take into account both the consequences and the states from which the consequences are generated. A causal relationship can thus be established between states and consequences. It would therefore seem reasonable to assume that a good description of the reasons for action is dependent on both states and consequences.[1]

However, this is not the most successful position in the social sciences and does not represent, so to speak, the common sense and general view. Instead, the widely accepted theory is the one that interprets,descriptively and normatively, the reasons for action as direct maximisation of expected utility. It is an elegant and intuitively simple theory, although analytically it is much less so than an intuitive glance would suggest, and probably its enormous success as an explanation of rationality in action must at least find one of its reasons in the simplicity of this position. This theory

[1] E. Anderson , *Value in Ethics and Economics,* Harvard University Press, Cambridge (Mass.), 1993.

found, as is well known, initial adequate support in the formalisation of von Neumann and Morgenstern.[2] It is an explanation of action that, by focusing only on the maximisation of desired consequences, leaves aside all consideration of ends. Only you, in fact, can be the judge of your ends, and thus the role that rationality plays in your choices can be nothing more than instrumental. 'Rationality' and 'instrumental rationality' thus become synonyms that can be substituted, *salva veritate,* in all relevant contexts – descriptive, predictive,normative – for the theory of action.[3] Presenting this view of action as a method for translating ought-to-be into being, however, may still leave room for a certain contamination with certain ends in the presence of other conditions and interpretations; for example, that which assumes an equivalence between expected utility and insurance rationality, and then makes it a formal characterisation of a mode of reasoning often informally present in liberal-democratic cooperative procedures, understood as a paradigm of rational cooperative action. This is famously the case with Rawls' philosophy.[4] Rawls' contractualism would not only duplicate the state of nature of classical contractualist theories, but, at least as, if not more, important, it would represent a kind of transcendental locus always available within the deliberative practices of our societies, a kind of paradigm of institutional procedures of how ought-to-be should be adjusted to being. Hence Rawls' claim that the theory of justice represents a part, perhaps the most important part, of a theory of rational choice. In both of these very different characterisations of the theory of expected utility as a theory of rationality – the one, avalutative, universal in scope, the other, cooperative, but no less ambitious it should be stressed a shift that is not without significance. This theory was first developed within the economic sciences as a predictive theory for highly characterised ideal situations. To embrace it – in one or other of its characterisations – as a theory of rational action and as a theory of correct reasoning for individual and social action is already to admit that the behaviour responds, perhaps not in detail, but at least in its essentials, to that primitive sphere in which the theory was elaborated.[5] Now, this may well be true and there are no a priori reasons to deny it, but at the same time there are no a priori reasons to accept it. What I propose in these pages is to examine whether the consequentialist paradigm behind this view. which specifies that theory, is an effective general description of our reasons for acting. The question is, ultimately, whether a unified rational theory of reasons and motivations for action is possible. This is a question that has a long and important history, of which I recall only two lines of development. One

[2] John von Neumann & O. Morgenstern, *Theory of Games and Economie Behavior ,* Princeton University Press, Princeton 1944; R.D. Luce & H. Raiffa, *Games and Decisions,* Wiley & Sons, New York, 1957.

[3] D. Gauthier , *Morals by Agreement,* Oxford University Press, New York, 1986.

[4] John Rawls, *A Theory of Justice*, Harvard University Press, 1971.

[5] P. Harnmond, *Consequentia/ist Foundations for Expected Utility Theory,* "Theory and Decision", 1988, pp. 25-78.

influential line of thought, which belongs to Hume, essentially denies even the possibility of this project. 'Reason is and must be the slave of the passions' is one of the mantras most frequently repeated by Humean critics. It should be noted that this statement could be one of the possible premises of an instrumental theory of rationality, which, however, does not at all arrive in Hume at a unified theory of reason in action, but affirms, behind a single name, an elastic plurality - a plurality traced back to the interweaving of the subjective circumstances of action, what Hume calls limited egoism and limited benevolence, and objective circumstances, the limited scarcity of goods. Another line of thought, for which Spinoza can be identified as one of the most influential mentors, identifies reason and knowledge. For this one, acting rationally means acting in accordance with the best available knowledge of the reality relevant to us. According to this tradition, therefore, a unified theory of reason in action seems possible as of now, although, at least in Spinoza's formulation, it seems to be so demanding that it can only become the patrimony of a few, although Spinoza also indicated socio-political correctives for this situation. However, even for Spinoza, acting according to reason, since it means pursuing for the agent in their own being, can be considered as the form of profit maximisation.

2. Is expected utility a good description of how we actually act and how we reason, and should a theory of reason in action be, normatively speaking, an instrumental theory? In order to answer this question, I think it is appropriate to try to quickly specify what forms an instrumental conception of reason in action has taken, even at the level of a certain common sense.

It is a view that holds several things together, since it has been configured both as a theory of behaviour and as a theory of correct thinking procedures in view of action. Actually, it is not entirely clear that a theory of behaviour can also be a valid description of thought processes. The reason, as I see it, is quickly stated: the idea of pursuing expected utility has, as a theory of thought processes, a markedly normative rather than descriptive dimension, in the sense that it prescribes which sets of preferences, desires, intentional attitudes, it would be rational to have in order to fulfil the requirements of the theory. It is doubtful that these sets can be determined a priori by reasoning that is not circular and tautological, that is, without the presence of some other set of assumptions. My conviction is that a broad set of knowledge must be assumed that the agent must possess rather stably, in the context of what, imprecisely, we may call a mature personality. This view, insofar as it deals with thought processes, has as its primary scope the strategies of individual action. Conversely, as a theory of behaviour, it has a much broader scope. It can cover the behaviour of political institutions, of trade unions, of occasional aggregations of

individuals, but can also be used to describe animal behaviour and sexual reproduction.[6]

Both of these characterisations of action are normative in two subtly different senses. What distinguishes them is that the first is normative ex ante, i.e. it undertakes to prescribe what is describable as rational behaviour through the evaluation of cardinally ordered alternatives; In the second case, on the other hand, normativity is often, so to speak, ex post, in the sense that behaviour that succeeds, gets its way, and has relevance to the majority of a group or to its most influential members and in a position to direct others, thus proves its rationality and is not instead assumed to be the result of what might turn out to be mere chance. This latter explanation also applies, for example, to interpretations in terms of fitn ess in evolutionary theories.

Besides being normative, are these two characterisations of action also descriptive and predictive? It is not, in fact, essential that they be so because there is no analytically necessary link between the success of an agent's action plans and the presence in their mind of 'correct mental processes' to achieve the outcome that guarantees the maximisation of expected utility. In particular, there is strong evidence to the contrary that shows that individuals are not able to correctly assess the probability of events that are indispensable for preparing the best action strategy. Successful prediction does not imply success in describing the agent's mental processes.

3. It remains true, however, that when acting, a certain instrumentality is always present, even if we are not prepared to reduce all relations between human beings to relations of economic exchange. This instrumentality can be described as the presence in the agent of a set of beliefs. Among these beliefs, at least three are considered to be particularly important: 1) the agent believes that an action is rational if it enables an end to be achieved that he has set himself; 2) reason determines the means appropriate to achieving the end; 3) the determination of ends concerns something other than reason, which acts as an instrument for determining means. As is well known, some go much further and think that reason can never determine ends, i.e. that reason is always instrumental. From this perspective, it therefore makes no sense to speak of a rational end, except insofar as it is instrumentally chosen, just as it makes little sense to speak of practical reason. Following some authors, one can, moreover, introduce a useful taxonomy that distinguishes between first-order ends (that which motivates action) and second-order ends (themselves means to the attainment of first-order ends). If I have a specific attraction to blonde women, I will not fill my house with posters of

[6] D. Satz & J. Ferejohn , *Rational Choice and Social Theory,* in "Journal of Philosophy", 1994, pp. 71-87.

Elisabetta Gregoraci; if I love small dogs, and do not own pets, I will not take a giant iguana into my home; if I wish to buy an apartment and have limited resources, I will not take out a mortgage to buy a Porsche (unless I am a victim of compulsions of some kind).[7] Note that an agent's action may retain all the characteristics of instrumental rationality, even if the outcome of the action is not at all the expected result, perhaps because adverse circumstances intervened or because, as is often the case, the information was incomplete. That is to say: the relevance of the process goes all the way to the decision-making procedure side of the action and not to its beginning - the desired end - or its conclusion - the outcome.

The agent's belief that he can achieve a certain end through a certain set of means does not, in fact, oblige him to assume the entire set of premises of instrumental reasoning, In fact, there are versions of rationality in action that accept the first two premises, but deny the third. Examples of this are various forms of Kantian deontologism or Nussbaum's capability theory.[8] In these theories, the end is not chosen by instrumental reason and, at the same time, it is denied that the choice of the end has no basis in other forms of rationality. In the Kantian versions, this is clearly visible in the notion of individuals as ends in themselves, i.e. as entities capable of envisaging the pure notion of duty as a guide to action, whereby in order to fulfil the moral law it is necessary to treat humanity in oneself, as well as in others, always as an end and never merely as a means. In Nussbaum's version, this is illustrated by the naturalisation of the theory of rights through the theory of capabilities. In these versions, instrumental reason is one of the manifestations, and not the most important, of a more general rationality. It is not, however, responsible for determining ends, not because these are extrational, but because their choice is a matter for a specific form of rationality. This form of rationality - directed towards the moral fulfilment of the realm of ends or towards the good life - discriminates between ends that are deemed rational and ends that are deemed irrational (depending on whether they correspond to some test of universalisability or the promotion of capabilities). These versions thus identify a more complex rationality that is more deeply committed to action, because its range of action is broader, encompassing the choice of ends and the determination of means appropriate to the nature of these ends.

This version of rationality is not an immediately predictive instrument, but since it possesses both a descriptive and a normative ascriptive dimension, it can be predictive in a mediated manner. In this sense, after the purpose has been ascertained and in the presence of relevant empirical information, one can also, say, argue that the choice of certain instrumental means may be erroneous.

[7] A. Mele, *Irrationality: An Essay on Akrasia, Self-Deception, and Self-Control,* Oxford University Press, Oxford, 1992.

[8] O. O'Neil, *Bounds of Justice,* Cambridge University Press, Cambridge 2000.

The idea that expected utility is the most coherent and simplest version of an instrumental reason is ultimately an idea that is not entirely clear because the predictive, descriptive, normative dimensions of utility are not clear. As a theory of reasoning it may find its place alongside or below a reason that also chooses ends; as a general theory of behaviour it appears as both a descriptive and a normative theory, and sometimes only in narrow fields.[9]

In this way, however, expected utility theory pays for its elegant simplicity with an illusion of totality, ignoring that norms are contextual just as ends are contextual, and that the means-ends distinction itself is a largely cultural product, varying according to the different vocabularies we use to account for our presence in a context relevant to us. Norms provide the agent with reasons to act, but the same instrumental reason may become an important part of the motivation to choose a certain course of action over another. Indeed, it may be that norms are an important part of the determination to follow a specific course of action or that they are even diriment to act in one direction rather than another, although claiming that an agent has reasons for doing a certain thing is different from saying that he is motivated to do that very thing.

4. Norms are of absolute importance in structuring our behaviour and can be variously binding and formalised. The norms for reading a newspaper, for example, prescribe that the newspaper should usually be read from the first page and not from the last. The rules for reading a newspaper, for example, prescribe that the newspaper should usually be read from the first page and not from the last. These are rules that can reasonably be assumed to have emerged partly spontaneously, as they are also generated as secondary effects of other rules - relating to writing, the possibilities offered by technical means, and so on. The rules governing offside in football, on the other hand, are more structured and less subject to variables - although these may be there and, indeed, play an important role in the final determination of the outcome of a match. Both of these classes of norms - examples could be multiplied at will, from road traffic norms to food etiquette norms - are cultural norms, whose motivational force can be explained using historical, sociological, psychological explanations. They are norms whose stipulative conventionality is explicitly established. There are, then, also other kinds of norms concerning convergence on focal points of behaviour, which have proved effective precisely from the point of view of expected utility, but adherence to which cannot simply be interpreted as the result of formal reasoning. There are, moreover, other norms whose culturality is not so evident and is, indeed, highly doubtful. The norm

[9] G. Carlson, *Plans, Expectations, and Act-Utilitarian Distrust,* in "Philosophical Studies", 1979, pp. 295-300.

that prima facie prohibits murder seems to be one of these.[10] The norm prohibiting the torture of children for fun seems to be another such norm. These are norms that appear to be transcultural, in the sense that in every possible world where we are allowed to imagine the presence of humankind these norms should apply. The idea here is that the presence of a natural kind - humankind - performs the authoritative function that is performed by social or psychological pressure or historical inertia in the previously mentioned contexts. In the latter case, the norms may indeed be violated or contemplate exceptions - even if there are no exceptions, other than pathological ones, to the norm prohibiting the torture of children for fun - such as when murder in self-defence is permitted or in war, but their prima facie authoritative force is in no way diminished by the exceptions. It is also in virtue of this consideration that some proclaim themselves objectivists in relation to the existence of these rules - which we can call 'values' - in the sense that being part of the human race means at least possessing mental dispositions to be motivated 'so and so' in the appropriate circumstances. In this sense, these are norms that have extra-cultural validity. I do not address the question of whether these norms express properties of the natural world, and, if so, how these properties can be ascertained. I merely note that those who deny that there are such properties on the grounds that we do not possess the appropriate epistemological tools to ascertain their existence should also argue that instrumental reason, for exactly the same reasons, cannot have the authoritative force that some are willing to grant it.

I believe that there is at least a possibility of understanding the normative force of instrumental reason without the need to make particularly demanding assumptions, in the sense that it does not seem to be particularly demanding to argue that on certain occasions the conclusions of instrumental reason may be persuasive enough to motivate on the basis of our psychology. It is, for instance, important to bear in mind that in certain decision-making processes a person may derive additional motivation to act from the fact that they have already behaved according to a certain motivational pattern in the past. These are incremental motivations, in which a causal connection is affirmed without assuming any precise necessary justificatory connection. There is nothing particularly challenging in arguing that, given our psychological construction, some forms of motivation are more attractive to us than others. In this sense, instrumental reason is motivational because it generates information for deciding between alternative courses of action.[11]

This position was, for example, held by Hume together with a moral dispositionalism that, combined with his sentimentalism, makes him a moderate

[10] J. Thompson, *The Realm of Rights,* Harvard University Press, Cambridge (Mass.), 1990; J. Bennen, *The Necessity of Moral Judgement,* in "Ethics", 1993, 103, pp. 458-472.

[11] T. Schelling, *The Strategy of Conflict,* Harvard University Press, Cambridge (Mass.), 1980.

proponent of a theory of instrumental reason. This kind of sentimentalism thinks of motivation as that which realises human dispositions, does not compromise itself with some form of invertible metaphysics, and accepts the challenge of the most comprehensive and rigorous explanation - the challenge of scientific explanation - by naturalising ethics. Hume is also the author who explains why utility appeals to us and is rightly considered a proto-utilitarian, but to embrace or found some form of utilitarianism that would prove adequate to what he considered to be characteristics of individual action and sociality he did not need any monistic theory of expected utility. I do not deny that his conception of reason in action is devoid of problems and unresolved knots. For instance, at a time when he accords considerable weight to reason, which ascertains what is within the reach of our desires, through logical reasoning and through causal ascertainment, should he not also naturalise this same faculty by making it a psycho-social mechanism?[12] Both scepticism about the self-substance and also sceptical notations about the original contract seem to lean in this direction.

If one can ask whether the kind of objectivity made possible by dispositionalism would not be lost because it would be included in a circularity with no external Archimedean points, it is necessary, however, to accompany this question with another: must we necessarily reduce normativity to non-normative elements? If the circle to which I have alluded is in any way subsistent, it would seem that it must be denied, since in imperative assertions what expresses normative direction - some verbal form of 'ought' - must be reduced to psychological expression with a great deal of caution and care.[13]

5. The success of formalisation among social scientists introduces two significant problems to a conception of utility that would naturally seem to have to do with desires and states of experience. The first is apparently a simple semantic shift. Utilities do not so much refer to desires or values, but rather to agent preferences. This entails a different approach to the problem of risk. This approach emphasises that the identification of utility with monetary values is completely ad hoc, since there are an infinite number of functions that increase in decreasing order, and undoubtedly the association with a monetary value varies from person to person, although it is not always entirely clear how. The second concerns the basis of this decision: why should a decision be based on the expected value of these utilities? It would seem that the only argument put forward for using the expected value is that it becomes apparent in the long run, when a game is repeated several times. There is merit in a frequency-based interpretation of expected value, but it is not at all clear why it should apply to an agent who participates in a game only once. Expressed to

[12] J.Broome, *Can a Humean be Moderate?,* in R. G. Frey &C. Morris (eds.), *Value, Welfare, Morality,* Cambridge University Press, Cambridge, 1993, pp. 51-73.

[13] A. Gibbard , *Wise Choices, Apt Feelings*, Harvard University Press, Cambridge (Mass.), 1992.

a certain approximation, what von Neumann and Morgenstem have shown for some is that if an agent is able to express their preferences relative to each possible pair of games constructed from a few simple alternatives and is guided solely by the value of their expected utility, then he/she acts in accordance with their actual preferences, provided there is a minimum element of consistency in their preferences. Once the agent has assigned numerical preferences to probable states of the world, we need nothing more to predict their scale of preferences than to verify that the axioms have been respected.

This theory is purely predictive in the sense that it predicts the agent's behaviour if the agent follows the ordering of preferences, an ordering that is shaped by the assumption of those axioms. It says nothing about what goes on in the minds of agents.[14] This cautionary assertion was also in accord with the general behaviourist cultural milieu within which von Neumann and Morgenstem provided their axiomatisation. However, mimalistic interpretations of the expected utility theorem are quite widespread, in the sense that it seemed consequential to assume that if the agent satisfies the axioms in the ordering of their preferences, he is at the same time an agent acting in accordance with the expected utility theory, as if an interpretation of the mathematical proof in a normative sense were occurring in their mind. Instead, I believe that what formalisation indicates is that it is possible to represent preferences and their ordering according to certain axioms; conversely, it is a misunderstanding to think that every representation of preferences and their ordering must follow that axiomatic ordering. This was a fortunate misunderstanding in the social sciences because it paved the way for the project of intra-personal and interpersonal utility comparisons. Against this misunderstanding, for instance, polemises Arrow, who does not believe at all that intrapersonal comparisons of preferences are possible in principle, and a similar polemical spirit can be found in those who deny the claim that the specific version of expected utility theory is the foundation of utilitarianism.[15]

Ultimately, expected utility theory is not at all a way of measuring preferences in order to order values and construct a representative image of what a rational agent should do. There is no indication in the demonstrative apparatus of how to establish analytical links with preferences insofar as they represent values or with preferences insofar as they rapresent welfare functions, although the idea that such a link is there or is at least implicit and easily deducible has found enthusiastic supporters among many non-economists. Rather, it shows how to represent preferences that order themselves in a particular way because they satisfy certain axioms, so that one can predict how an agent should choose under conditions of risk. The conclusion that

[14] P. Hammond, *Consequentialism and the Indipendence Axiom*, in B.R. Munier (ed.), *Decision and Rationality*, Reidd, Dordrecht 1988; A. Sen, *Evaluator Relativity and Consequental Evaluation*, "Philosophy and Public Affairs", 1983, pp. 113-132.

[15] K. Arrow, *Social Choice and Invidual Values*, Wiley, New York 1951.

because the proof is valid, then agents are utility maximisers is undue, as is this other: that agents' preferences are generated by some deep psychological structure that is formalised by the proof. The numbers generated by a von Neumann and Morgenstem function reveal nothing about 'utility in general' and nothing about the measure of 'well-being'.[16] A link between the utility function generated by the adherence of a certain representation of the agent's expected utility and certain values should provide circumstantial evidence of the necessity of this link, yet the knowledge of the necessity of this evidence is quite scarce. This link is unlikely ever to be proven in an analytical sense, because expected utility theory is normative, and does not provide us at all with a predictive model on the basis of which to judge human behaviour. This means that nothing entitles us to think that, assuming that our preferences defy the initial axioms, there are mathematical functions in our minds that guide and direct our behaviour, or even just that our behaviour should conform to these functions when we make choices. It is, in fact, a tool to represent an order, and not even the only one. It is not, however, a theory of reason, or a theory of reasoning in action.

6. It was said that there are also reasons related to the cultural temperament of the time in which the theory was developed, which support an interpretation of expected utility as a specific and not unique representation and not as a psychological function. Indeed, the idea that expected utility represents internal states in the minds of subjects would not have passed any of the neo-positivist and behaviourist tests of the relevance and significance of how we evaluate something through the utility function, as this evaluation is something distinct from a simple disposition to choose.

When one interprets expected utility as a model of behaviour that is generated by our mind, the operation one performs can be of twofold nature: either one is interpreting norms of behaviour or one is interpreting reasoning processes related to action. By adopting either of these two interpretations - sometimes both - the least that can be said is that the theory is being forced in the direction of an ethical-practical result - 'one must act in this way' - or a metaphysical gnoseological one - 'these are some of the characteristics of the mind' - neither of which are found in the axioms. But up to here, nothing wrong, of course: if it is considered a legitimate hermeneutic activity to think what a thinker did not actually think, but should have thought on the basis of their premises, then there is no reason to consider any extension of a theory into fields, which did not originally enter into its initial formulation, a priori a misunderstanding. However, this extension of expected utility towards a cognitive and structural dimension is at least in part a

[16] G. Loome & R. Sugden, *Regret Theory: An Alternative Theory of Rational Choice under Uncertaincy,* "Economie Journa1", 1982, pp. 805-824.

misunderstanding, though there is no lack of reasons for it. I would point out at least one, namely the idea that the ordering of preferences that should guide certain choices refers to a set of psychological laws that are identical for all human beings. But is it necessary for a purely predictive theory to be based on an assumption that is as demanding as it is unproven? Is it not true that the subjective - and cultural - propensity to risk is a fact that it seems at least undue to exclude from the search for utility? The same difficulties about the collection of relevant information - about the difficulty of deciding which information is relevant and which is not - might suggest, to be cautious, that many agents do not act only after attributing defined intersubjective probabilistic values to events, even without concluding that they behave irrationally.

If the theory of expected utility was not intended to function as a theory of reason or as a theory of correct reasoning, it could be argued, however, that adding normative and psychological corollaries is a legitimate and indispensable operation for the completeness of the theory itself. For this reason, placing the psychological superstructure, which considers the agent's preferences as what defines the ends of its action, alongside the formal interpretation should provide a much less indefinite sense of the whole affair, since it is concluded that not every preference can perform this function, but only those that satisfy the axioms. This point is decisive, because if there is a normative sifting through which our preferences must pass in order to be usefully considered as end-formers, this means that the axioms are not instrumental and that therefore the theory of expected utility itself, insofar as it is nothing more than a derivation from these theorems, would have a normative scope. Also, therefore, using the forms of reasoning, which are based on these theorems, would have a normative scope and not instead an instrumental one. This position, for example, was taken up by David Gauthier, who considered the axiom of transitivity as a kind of test of the consistency of our preferences, thus interpreting it in a normative sense. The next step is to read axioms as rules of logic, thus suggesting that they are non-instrumental rules that constrain the consistency of a given set of our preferences. All this seems to violate the Humean mantra that reason is only a slave to the passions - even though Hume said that reason is and must be a slave to the passions, appropriately mixing the descriptive and prescriptive levels -. Once we have made this small deviation from Hume's groove, how can we be sure that other normative principles should not be invoked? How can one be sure that, for example, ethical-moral considerations should not be brought into play to ensure some form of stability over time for our preferences? We agree that stability over time is not the same thing as logical consistency, but why should the former be less important than the latter? The stability of cooperation is in fact based on an arrangement of preferences of each of the agents involved, which is assumed not to change over a relevant period of time. The tit/or tat strategy is precisely a bet on this kind of stability. This stability is at least as important as the formal

consistency of each agent's preferences. The same applies to other normative standards that might be called into question: for example, standards of an aesthetic or religious nature. This means that, since a non-instrumental, i.e. normative conception of reason underlies an instrumental conception of reason, the whole of this view is ultimately non-instrumental, unless one believes that the axioms that are supposed to shape the set of our preferences at stake are so obvious that they do not require any discussion. This, however, is far from obvious.

There are various strategies to avoid this mixing of the normative with the instrumental that we must now briefly examine. A first move is this: to argue that while it is not immediately obvious that axioms are instrumental constraints on our preferences, it would remain true that we have instrumental reasons for following them. That is to say, since the result of the shaping of our preferences, which are structured through the axioms, is consistency, one could say that an agent who wishes to achieve their ends by maximising their utility would do better to be consistent in their preferences. This is advice and a prescription that results in a consequentialist defence of the axioms. In other words: it is a defence that says nothing about the nature of those axioms and, in particular, nothing about whether or not those axioms represent a formalisation of natural and fundamental rules in the reasoning of the human mind. To this it must be added that this position, on the other hand, does not seem to suggest a coherentist defence of the appropriateness of adopting the system of axioms, since it does not say that we have good reason to accept them since they are shown to be in agreement with other principles of the mind. What is being said is that experience seems to suggest that adopting a coherent arrangement of one's preferences with feels relatively more direct - with an adverb and adjective also imbued with normativity - attainment of one's goals. It follows that when one says that an agent has a set of preferences that is not consistently structured, because the order of this set does not allow him to maximise utility in the standard version of the theory, what one is saying is actually a tautology, since expected utility is defined in terms of structuring one's preferences according to those axioms.

It could be argued, however, that if the theory does not specify the precise way in which the means to a specific end are determined, this is by no means a good reason to deny the intuitive basis of the theory, precisely that which was referred to at the beginning of these pages, namely, that what the agent does is what he believes should actually be done to satisfy a set of desires. Moreover, precisely such an intuitive consideration would seem to underlie the emphasis on the consistency of preferences among themselves. But this brings us back to the point made above: it is not so much instrumentally rational to follow axioms in order to satisfy one's preferences; rather, it is the use of these axioms that makes the set of preferences coherent, and thus satisfying these preferences is part of what instrumental reasoning and behaviour consists of. It may well be that some perspective on the

means-end correlation is necessary, just as it may well be that a comprehensive consideration of the reasoning that determines the means in view of these ends is necessary; however, this is manifestly a different thing from the characterisation of theory as a rational structure that prescinds from value considerations. If axioms provide a kind of sieve through which our preferences must pass before we can claim to have acted rationally on the basis of those same preferences, this means that they define what it means to be an end in an indirect way. So the conclusion to be drawn from this is that these axioms cannot be defended on the basis of consequentialist considerations.

Furthermore, there is a great deal of empirical evidence showing that real agents continually violate the axioms of the theory. Some of these violations would seem intuitively reasonable, so that the classical expected utility theory would not even seem to be able to be proposed as a plausible theory of human behaviour. I will limit myself to two axioms of the theory, the axiom of transitivity and the axiom of reduction of compound lotteries.[17]

The axiom of transitivity states that if an agent prefers a to b and b to e, then it also prefers a to c. Let us now imagine an agent who can perform three different actions. Each of these three actions has three possible outcomes, all three of which are equiprobable. We can imagine a perfectly consistent situation where a is preferred to b and b is preferred to e, but where e is preferred to a. If I prefer coffee (a) to chocolate (b) and chocolate (b) to tea (e), I can in all consistency prefer tea (e) to coffee (a).

The assumption behind the compound lottery theorem is that if it is possible to reduce a compound lottery to a simple lottery, the agent will be indifferent to the different lotteries. The significance of this axiom is doubtful, because it is not obvious what its relevance is, and some doubt that it can really be considered an axiom, but requires further assumptions. In fact, human behaviour seems to be a rather frequent violation of this axiom. As a rule, individuals are not indifferent to the number of steps required to determine the outcomes of their actions. If I have a desire to buy a new car and I adopt the solution of applying for a loan instead of saving for four years, I certainly cannot say that the latter solution is equivalent to the former. The desire to avoid certain intermediate steps - four years of savings - may lead me to prefer one resolution over another.

One can also give the case of a person who is sensitive to a lottery that contemplates several intermediate outcomes towards the final outcome and is content to reach some of these outcomes, even if reaching them does not increase the probability of reaching the final outcome. This seems to be the experience of gamblers. If I offer a sum of money to a hardened gambler, a sum he hopes to earn

[17] D. Kahneman, P. Slovic, A. Tversky (eds.), *Judgment under Uncertainty: Heuristics and Biases,* Cambridge University Press, Cambridge, 1982.

by playing, provided he does not play, what will happen? He will most likely behave like Dona Flor's husband in Jorge Amado's novel. He will give the most ample assurances that he will not go gambling and, having pocketed the sum, he will head for the nearest gambling den. This is not just a psychological notation, as much as it signals the fact that it is important to the gambler how he achieves the result he desires. This is a similar experience to that experienced by agents wishing to achieve a result, which requires lengthy professional preparation. There are, of course, cases where the professional summit is reached early in one's career in a fully deserved manner, but these are generally rather rare cases. There are, on the other hand, cases in which this summit is reached quickly due to circumstances extrinsic to the profession (friendships, political support, and the like). Many people would reject such a rapid resolution of their careers in the absence of circumstances based on merit. The reason for this is quite clear: it is part and parcel of the lottery's utility that one prefers it to play out in a certain way.

Some agents like risk and suspense and will place a different and higher value on a lottery, which contemplates different stages of anxiety towards the attainment of the final goal, than on a simple lottery where this is reduced. There is no obvious reason to think that an attitude towards risk cannot be contemplated in a calculation of utility for the agent, because the manner in which a lottery takes place may be of entirely relevant value to him. That is why a lottery with more steps may be preferred to a simpler one. A descriptive account of human practice in risk situations would be incomplete if it ignored this possibility.[18]

In many lotteries, the pleasure of playing one way, rather than another, must be included in the same utility function. If I want to win a person's attraction and have the choice of either inviting them out to dinner in the hope of starting a formal relationship or having them unwittingly ingest a psychotropic substance that will make them fall in love with me, it is not irrelevant how I achieve my goal. In fact, there is no good normative reason to deny a legitimate propensity to risk.

In other words, the axiom does not allow for a positive attitude of the player towards the very fact of participating in a lottery. It is, however, a moralistic aversion, which denies that the consideration of direct sources of utility to the very fact of playing a particular game, for example, based on some lexical graphic order, can come into play in the agent's choices. The idea that preferences are only on the results of lotteries and cannot also be on the quality of lotteries seems, therefore, to be very counterintuitive.

7. Preferences can be very diverse. I may want my immediate pleasure, the satisfaction of a particularly nice neighbour, world peace, the reduction of international debt, and so on. In this sense, the utility function is supposed to be a tool for measuring and ordering preferences in the mind of the agent, but it is also

[18] D. Kreps, *Notes on a Theory of Choice,* Westview Press, Boulder 1988.

not a tool that evaluates and decides what the preferences are for. This interpretation of the utility function is not even strictly speaking a theory of reasoning. It simply notes that the agent has preferences whose motivational sources may be very different and that these preferences may be represented in a certain way. Arguably, this weaker interpretation is the one favoured by most theorists because it seems to be particularly parsimonious in its assumptions, whereas Bentham's classical monistic interpretation is much more problematic, insisting that, ultimately, we have only one preference in our mental structures, the preference for pleasure (for its overall maximisation), so that if some preferences appear to be inconsistent with each other, the representation allowed by the application of the axioms ensures that the overall result expected from the reformed preference order is not. In both interpretations, the pluralistic and the monistic, the axioms would thus be a way of generalising a mode of reasoning that informally exists in the psychological structure of human beings, the formalisation of a mental function if not, indeed, of a biological function. Even this portrayal exposes itself to nonconsequentialist considerations insofar as the mapping of preferences that the agent draws is, at least in part, a function of the world in which consequences are expected to take place, as the result of lotteries, of states of certainty, of what he interprets as the very context in which the choice takes place. For this reason, the psychology of the decision maker must be a more complex psychology than that envisaged by expected utility. Unless one believes that preferences are generated in us in such a way as to be spontaneously consistent, the axioms of the theory cannot but have a normative aspect. From this point of view, their function should be more to make us aware that we cannot really want contradictory things, than to describe a way of reasoning. The idea that the agent must be able to distinguish not only its preferences with regard to outcomes, but also its preferences with regard to the way in which outcomes are generated, cannot be set aside at all, unless an axiom is purposely introduced.

Let us assume that you have to choose in a lottery between 200 euros certain and a certain sum S uncertain, but greater than 200 euros. While the 200 euros will not increase your marginal utility, if only slightly, the winning of S could, perhaps, change your life. You can argue your preference for the 200 euros by claiming that you do not like risks, that lotteries with uncertain outcomes, even if extremely profitable, make you nervous and cause you suffering that can only be alleviated by knowing the outcome, even if it is unfavourable to you. It is, in this example, the very context of the choice that determines your preference.

Let us now imagine that an agent says he prefers S because he is inclined to take risks and because the expected utility of S is higher than the utility represented by the certainty of 200 euros. In this version, the utility of S is not at all the same as in the first version, because in the first example S also incorporates for the prudent (perhaps over-cautious) agent the disutility of procuring the stress associated with

participating in the lottery itself. For this reason, the utility of S in the second situation is apparently higher than the utility of S in the first, because in the latter we have traced a certain feeling of risk-aversion on the part of the agent.

In each choice, one can find, by investigating thoroughly enough, motivations that can be traced back to slightly different contexts of the choice, which make seemingly similar situations, in reality, different. This can be embarrassing, because it could justify any agent from being accused of violating any theorem. Since differences between even extremely similar alternatives can always be traced with a little good will, then any accusation of irrationality to any action could prove unfounded and essentially irrefutable. Such a theory, since it continually resorts to ad hoc justifications, seems useless as a descriptive tool of action and as an instrument of prediction; and yet, this does not entitle one to infer that there are good normative reasons for excluding the context of choice.

Some have argued that this is actually not a problematic situation at all. All we need is to observe how the agent behaves in a situation of uncertainty, i.e. how he openly manifests their preferences. This prescription, however, does not tell us at all how to discriminate a preference from a state of indifference. If I go into the greengrocer's, try to look around and then come out with a kilo of apples, this does not at all inform whether I simply preferred to buy something in the first shop I came across on my way, whether I intended to buy a kilo of peaches but then fell back on one of apples because peaches were too expensive, or whether I prefer apples to peaches. This uncertainty occurs due to the simple fact that the activity of preferring is an interpretative activity and not a simple observational event available to everyone 'out there in the world'. What interpretation then should we make use of to correctly identify preferences and their order? One way might be this: identify the alternatives through which preferences are formed. Some of these will be indifferent to the agent, i.e. preferring one or the other will change essentially nothing in its expected utility. This procedure is given the name of the rational indifference principle. Note, however, that even this principle is, once again, markedly normative, since it tells us what an agent should do when faced with two reasonably similar situations on the basis of some principle of individuation, i.e. he is essentially told, again, what preferences he should have in given situations, resorting to the intuition of two reasonably similar situations, without taking into account that the idea of 'reasonably similar situations' can be powerfully influenced by cultural pre-suppositions and individual differences. This makes an important part of utility theory dependent on some uncertain appeal to intuition. However, perhaps it is not necessary to embark on the path suggested by the principle of rational indifference and accept quite simply that preferences are subjectively defined, and that this definition is part of the psychological states of the agent. These preferences are identifiable because they are the object of the agent's experience. In this sense, if remorse or fear or joy or elation are part of the agent's expected

consequences of the experience of an action then we should include one, some, all of these feelings in the calculation of utility. For this perspective, the course of action the agent chooses is a fact of reality from which one cannot abstract any more than one can abstract from the feelings associated with choosing one particular course of action in preference to another.

What might this suggest? That it does not seem enlightening to incorporate preferences for states into preferences for consequences, when in fact the preferences are for states and not for consequences. The agent can act independently of the probability information he has acquired about the consequences of their actions, favouring states of the world in which different compound lotteries take place that remain irreducible for him.

What is the moral we can draw from these last reflections? That the expected utility theory in its consequentialist interpretation as a tout court theory of reason is inadequate. Our reason, our reasoning is often sensitive to other things than consequences. This is a simple fact that we derive from experience and there is no descriptive or prescriptive reason to underestimate it and not take it into account. Of course, none of this means at all that expected utility theory is not useful in economic contexts for describing behaviours that occur in the market, or even for arguing that these specific behaviours are perhaps the most relevant in that context. Economists, therefore, have every reason to use theory to make predictions. The point is not to confuse the ability to make predictions in given contexts (assuming this is possible, which it is certainly not within my competence to judge) with some monistic theory of the capacities, inclinations, propensities of the mind.

If the theory is really predictive in certain narrow fields or, more likely, in certain circumstances within these fields, this is not because it is a theory capable of providing an ade guate image of the way in which we reason or the way in which we should reason. Now, since there are forms of reasoning and behaviour that reveal violations of the axioms of the theory of expected utility, and which do not appear to be irrational at all, if modifications are necessary, they do not appear to be on the side of behaviour and action and its justifications, but on the side of theory. The justifications for action will have to be situated in a context that takes account of the fact that not all our acting is consequentialist, that is, that often our reasoning is contextual in a deep and non-episodic sense. If consequentialism is a falsifiable theory and not some kind of mythical theory of acting, the idea of contextual dependence should be considered as a kind of test of the conceptions of reasoning on which it is based. Evidence that each of us can draw from our own lives, but which also comes to us from the empirical sciences, strongly suggests that we are not only consequentialist agents, although we are also agents who operate on the basis of assumptions about consequences. Can this pave the way for a theory of behaviour that integrates the axioms of expected utility theory while retaining its instrumental depth? I think one must be highly sceptical of such a perspective. The

motivational sources of human action are manifold, and reductionist moves, although by no means illegitimate a priori - and are, moreover, historically recurrent - as well as arguably necessary for modelling purposes, miss the essential of human behaviour, namely its variety in action and in the sources of motivation. But there is another consequence that I think can be drawn, and which is of greater relevance to moral philosophy, and it is this: if our reasoning for action has important features that are not subsumed under the consequentialist module, why should we think that our moral action is? As much as believing that consequentialism and instrumentalism are synonymous may be a mistake, and a consequentialist agent is not bound by the idea that their actions are all instrumental, such an agent remains, nevertheless, bound by the idea that the balancing and choice of actions to be performed are framed in the consequentialist module. But this module can take a plural variety of forms and, therefore, an instrumentalist perspective implies consequentialism, but the reverse is not true.

Preferences for states and preferences for consequences are fundamentally different and cannot be reduced to one another, which is equivalent to saying that there is by no means a single kind of 'thing' that requires to be maximised, nor a single procedure for doing so, nor a single procedure for maximising different things; it follows that prescriptions on how to satisfy a set of preferences cannot at all avoid taking into account the integration and mutual modification of different kinds of preferences. This necessity is, I think, particularly clear if we imagine that some preferences for states are moral or have a moral basis.

Having said that, could it still be argued that expected utility theory is a form of idealisation, which like all modelling leaves something aside to focus on what is deemed most relevant? I think it has to be said that while we certainly need modelling, we need models that conform to the object. If practical action does not conform to the model of expected utility, there is clearly no point in saying that it is action that needs to be reformed. Instead, it is the need to formulate a better model. A misrepresentation of motives in action is particularly unfortunate in the normative sphere, both because it surreptitiously introduces the idea of an idealised normativity that would be derived from scientific theories and because it fails to recognise that, if normativity is to be there, it must also be situated on the side of principles governing the integration of preferences. It is possible that a defence of this perspective will have to assume a coherentist strategy. This strategy will have to balance both our expectations of consequences and our preferences for states. It will probably still be unacceptable to those who cultivate dreams of social engineering, but in its inevitable approximations it will be closer to the pluralism of human action.