

Julian Nida-Rümelin

Why Rational Deontological Action Optimizes Subjective Value

Abstract

In present day philosophy there are two competing views regarding practical rationality: (1) Decision and Game theory and economic theory have developed a theory of rational decision which has proven to be fruitful in many areas of social science. Practical Philosophy should work with that paradigm (2) Economic theory and decision theory do not have an adequate account of practical rationality. The *homo oeconomicus* model is – at best – one perspective which competes inter alia with philosophical accounts of practical reason.

In this article I try to show that these two seemingly opposing views are in fact compatible. I argue that consequentialism is an inadequate account of rationality because rational action is deontological in character. Nevertheless the decision theoretic conceptual frame should not be given up. Deontology and decision theory can be made compatible via *comprehensive description* of action. The conceptual frame of decision theory should be interpreted as coherentist, not consequentialist. With this interpretation deontological action, if rational, maximizes subjective value.

I. Introduction

Does rational deontological action optimize subjective value? Can one be a deontologist and at the same time adhere to decision theory as an all-embracing theory of practical rationality? I think the answer to both of these questions is *yes*. The reasons for it are given in this article.

There are two basic intuitions which frame the bigger part of practical philosophy and which seem to be incompatible. One intuition is *teleological* or more specifically *consequentialist* according to which rational action optimizes its consequences. The other intuition is *deontological* or *rule-oriented* according to which rational action is guided by certain rules. I am a deontologist, I think that consequentialism is an inadequate theory of ethics and rationality alike, but at the same time I am convinced that rational action maximizes subjective value.

The reader probably thinks that the following two assumptions cannot be true simultaneously

(1) Consequentialism as a theory of *rationality* is false

(2) Rational action maximizes subjective value

respectively

(1') Consequentialism as a theory of *morality* is false

(2') Moral action maximizes subjective value.

I try to show in this article that this alleged incompatibility does not exist. The teleological intuition can be upheld not in the original *consequentialist*, but in a weaker *coherentist* form. A deontological (anti-consequentialist) theory of rationality and morality, as I think it to be adequate, is compatible with coherent preferences and

beliefs and fits therefore into the decision theoretic conceptual frame, if interpreted in a coherentist manner.

If one takes moral actions to be rational actions, there is a close link between ethics and the theory of practical rationality¹. The best-developed theory of rationality, however, is decision theory (including game theory), and decision theory is a consequentialist theory of rationality, or so it seems. Most ethical deontologists share the common philosophical belief that decision theory is about prudence (or strategic action), whereas morality establishes constraints on prudential optimizing. Deontologists, therefore, tend to be hostile towards the idea of taking decision theory to be the core of a general theory of practical rationality and towards the attempt to integrate moral actions into its conceptual framework.

If the decision theoretic framework cannot be dismissed in order to define rational action, it seems that ethical consequentialism has strong arguments in its favour. Ethical axiology would be about moral values and ethical rationality about how to optimize moral values. Utilitarianism and decision theory which both are closely interlinked in the history of ideas would then still be dependent on each other from a systematic perspective, too. I will instead argue that rational deontological action is compatible with standard decision theoretic axioms. *Rational deontological action optimizes subjective value.*

II. The Wedge between Choice and Preference

¹ I developed the arguments in favour of a deontological account of rationality and morality in *Kritik des Konsequentialismus*, München/Wien, second ed. 1995, in *Strukturelle Rationalität*, Stuttgart: Reclam 2001 and in *Decision Theory and Ethics*, München: Utz 2005, therefore I shall not repeat these arguments here.

Decision theory works with the assumption that the rational person reveals two basic propositional attitudes in acting: *subjective preferences* and *subjective probabilities*. It attributes two real-valued functions to the rational actor: the function of *subjective probability* and the function of *subjective value*. This attribution is based on the comparative concepts *preferring* and *expecting* - person *i* *prefers p to q* and *i takes p to be more probable than q*. It is not possible to attribute one of these two functions independently from the other, and insofar the attribution of the two functions is interlinked.

Is decision theoretic rationality consequentialist? The standard applications of decision theory are indeed consequentialist insofar as *subjective value* is defined as *subjective value of consequences*, and insofar *rationality* is defined as *maximizing expected subjective value of consequences*, or in short as: *optimizing consequences*. Conceptually this consequentialism is made explicit in Savage's model, for example, but not in Jeffrey's holistic account². If we think that moral actions should be rational actions, morality should be defined as maximizing expected subjective value of a *specific kind*. It would be the task of *ethical theory* to develop criteria under which a subjective value function seems morally acceptable.

What, then, if you are a deontologist, that is, if you think that deontological criteria cannot be dismissed in ethical theory? Under the assumptions that the decision theoretic framework can not be dismissed either and that decision theory commits one to a consequentialist view on rational agency, the application of deontological criteria seems to suggest the following options: either the idea that moral action is rational has to be abandoned, or two different kinds of rationality have to be adopted: one for moral and

² cf. Leonard J. Savage, *The Foundations of Statistics*, New York 1954 and Richard C. Jeffrey, *The Logic of Decision*, New York 1965.

one for extra-moral action - in the latter case, decision theory would have to be taken to deal with extra-moral actions only.

The first option appears unattractive: "You should do x, but it is irrational to do x!" is not a convincing moral imperative. The second option would imply that decision theory is not a general theory of rational action. In this article, I shall not be concerned with the first option (we dismiss it as implausible without further argument), but focus on the second: Are we to accept that there are two quite different kinds of rationality? Is it true that deontological action does not fit into the decision theoretic frame?

Decision theory does not begin with postulating probability and value functions. Rather decision theory begins with describing *basic properties* of preference relations. Some of these properties can be interpreted as analytically tied to the notion of preference, others might be understood as synthetic (and normative insofar as they, taken together, constitute rationality or are at least necessary conditions of rationality.). Here - as elsewhere - there seems to be no clear-cut borderline between *analytic* and *synthetic*. On the one hand, it seems that *reflexivity* of preferences is an analytic property, whereas *transitivity* of preferences rather appears to be a synthetic (and normative) requirement. On the other hand: if a person has preferences regarding three well-determined, well-described and well-known alternatives which are intransitive, one may well ask the person, whether she *really* has those preferences. If we think of preferences as being revealed in action and in verbal expression, violations of transitivity or other properties of rational preference relations may result in uncertainty concerning the reliability of the attribution of these preferences in the first place. This phenomenon indicates that there is no precise demarcation line between analytic and synthetic properties of rational preference relations.

Is there any reason to assume that deontological action violates one of these properties of rational preferences (whatever their status may be), e.g. transitivity? This formulation is an elliptic version of: is there any reason to assume that actions which are performed for deontological reasons reveal preferences which violate the transitivity condition? It makes things easier if we circumvent the notion of *action* by asking: is there any reason to assume that *preferences* which are based on deontological rules violate the transitivity condition?

Let us illustrate the point by referring to the deontological rule *THOU SHALT NOT STEAL*. The conflict between deontology and consequentialism is apparent: Even if stealing in some cases might have overall good consequences, you are not allowed to steal. This remains true also if we concede that in cases of extremely bad consequences e.g. starving, stealing might be justified.

Now think of the following case. There are three possible states of the world x, y, z . We assume that these three states of the world are certain³ consequences of three alternative feasible actions u, v, w respectively. The person in question prefers x to y and y to z . The subjective value function represents these preferences adequately. Since probability is not involved, maximization of expected subjective value would result in choosing action u . But let us assume that u is an act of stealing, whereas v and w are no acts of stealing (and are not forbidden by other deontological rules). In this case it seems that v should be the preferred action. The situation is such that it is not possible to optimize consequences and at the same time conform with the deontological rule in question.

³ "Certain" in the sense that the subjective probability is 1.

We can assume that the rational person has *transitive* (and in general *coherent*) preferences regarding states of the world as well as regarding actions. Both kind of preferences should be coherent, if the person is assumed to be rational. But the rational person's preferences regarding states of the world do not necessarily determine the person's preferences regarding actions. This is not implied by the axioms of the utility theorem. It is implied only by specific interpretations, that go beyond what is required by the axioms themselves.

Reflexivity of weak preferences is given trivially (in our examplee there are no weak preferences involved) and *Completeness* of preferences is a reasonable postulate for a deontologist and a consequentialist alike. Since the rational deontologist (like the consequentialist) cannot refuse to act, she reveals her preferences regarding any feasible alternative. There is no reason to assume that a rational person with deontological motivations gets sometimes caught in holes of inactivity. The *Completeness Axiom* is valid for preferences motivated by deontological reasons, too.

One widespread form of presenting decision theory identifies actions with prospects, in the sense of probability distributions over the sets of their consequences and takes consequences as states of the world (or sets of states of the world). Therefore it identifies u with x , v with y and w with z . Within this conceptual framework there, indeed, is no logical space for deontological rationality. It is excluded by conceptual means that a rational person has preferences as we assumed above. *This conceptual exclusion of deontological rationality is unacceptable for any adequate theory of practical reasons and it is unnecessary in order to preserve the conceptual frame of decision theory.*

It makes perfect sense that a person, when asked which state of the world she prefers, answers that she prefers u to v and v to w while she respects the Seventh Commandment and therefore prefers action y to action x . There is no reason to think that everybody, who respects this commandment is irrational. A rational person can prefer the state of the world x to y and y to z and at the same time prefer the action v to the action u , because u would violate the Seventh Commandment and v wouldn't.

With this example in mind, the flaw in standard consequentialist decision theory is apparent: Preferences regarding states of the world are insufficient to determine rational action. *We have to drive a wedge between choice (of actions) and preference (between states of the world)*. The adequate reaction to this flaw is to avoid any conceptual dependency on consequentialism. Even the friends of consequentialism cannot wish that consequentialism is true by definition. If consequentialism is true it is true by synthetic normative argument. If it is wrong, it is wrong by synthetic normative argument. If the *open question argument* is of any value at all, it can well be applied to this case: "Action x has u as its consequence and u is the best of the three possible consequences u, v, w , but is x rational for somebody who accepts the Fifth Commandment?" is a reasonable question, that cannot be ruled out by conceptual means.

We are therefore forced to draw a conclusion that has far reaching implications for any adequate theory of practical rationality: *Not all actions can be represented by their prospects* in the sense of probability distributions over the set of their consequences, if "consequences" are understood as "states of the world" or "sets of states of the world". Actions change probabilities of propositions (understood as sets of states of the world), this is how actions intervene in the course of history. The prospects of an action can be quite relevant to determine whether it is rational. It is, however, an essential trait of

consequentialist persons only that for them actions can be represented by their prospects (understood as described above). But the preferences of a rational person regarding prospects (sets of state of the world) *can* differ from the same person's preferences regarding the actions that cause these prospects. *It is not universally true that one can substitute actions for prospects within a rational person's preferences.* It is not true in case of *deontological* rational actions. *Deontological actions cannot be represented by their prospects in a rational persons' preference relation.* The rational and deontologically motivated person drives a wedge between choice (of actions) and preference (of states).

Deontological actions are motivated by deontological reasons. An action complies with deontological reasons *under a description that makes this compliance apparent.* But this is exactly why deontological actions cannot be represented by their prospects. To describe an action by its prospects alone excludes that under this description it is rationally motivated by deontological reasons. An action can comply with deontological reasons only under a *comprehensive* description as opposed to a *reduced* description by sole reference to its prospect.

III Ramsey-Coherence

The question remains whether it makes any sense to apply the decision-theoretic framework to *comprehensively described rational deontological actions.* In order to answer this question we have to check whether the basic axioms of decision-theoretic rationality are applicable to preferences regarding comprehensively described rational deontological actions. If u, v and w are three comprehensively described rational deontological actions, a rational person who prefers u to v and v to w prefers u to w .

This requirement of *Transitivity* should be met not only in case the rational person's preferences are based on consequentialist, but also if they are based on deontological reasons. Why should deontological reasons go against transitivity? Why should a rational agent develop intransitive preferences only because he is motivated by deontological reasons? Why should deontological motivations lead to intransitivity?

I think that there are no reasonable answers to these questions. Moreover, if a set of deontological reasons justified intransitive preferences this set of deontological reasons *ipso facto* would be shown to be incoherent: Transitivity of preferences is a requirement of practical coherence that can not be dismissed if we move from consequentialist reasons to deontological ones. Likewise we can assume that the decision-theoretic requirements of *Reflexivity* and *Completeness* are met by the rational person's preferences within this set of comprehensively described rational deontological actions.

Problems seem to arise if we include probabilistic considerations. We are used to interpret prospects that are assigned to actions in a consequentialist manner. In doing *a* I expect the consequence *c* with probability 0.2 and the consequence *c'* with probability 0.8. But what means "consequence" here? The verbal expressions we use for describing consequences are underdetermined regarding this question, because the answer is dependent on what we take to be relevant for a rational preferential ranking of actions. It might be understood as a 'consequence' of doing *a* that I keep my promise or that I obey the Seventh Commandment, such a 'consequence' though is not a state of the world or a set of states of the world, in a consequentialist sense. It is a proposition, but a proposition that can be true or false without changing the state of the world. It might in this sense not have any 'consequences'. If you take 'consequence' as being – more

specifically - characterized by the state of the person's mind⁴, this remains true a fortiori. The decision-theoretic axioms, however, should not determine what is relevant for establishing a rational preference-relation. They merely should constrain the form of it in order to secure practical coherence, they require *inter alia* that the preference-relation is an ordering.

As explained in the last section deontologically motivated preferences require *comprehensively described* actions. The reason is that the prospects of actions in the usual consequentialist sense do not include enough information for the deontological agent to rank his actions. But if this is true it seems that we lose any ground for including probabilistic considerations. And if probabilistic considerations were excluded we could not apply decision-theoretic axioms that take probabilistic considerations into account (the monotonicity- and continuity-axiom). In this case the decision-theoretic framework would collapse. It would become impossible to assign a well-determined (up to positive linear transformation) function of subjective value to the rational deontologist. The transformation of preference into value would not succeed any more.

Luckily there is a way out. In order to find it let us remember what Elisabeth Anscombe⁵ and many philosophers of action after her described. Actions result in other actions and it is often not clear to the actor in which actions his actions result. If I press the switch it is not sure whether this action results in the action to switch on the light, because there might be a defect in the electrical wire system. If I switch on the light by moving the switch, it may be that I frighten my wife by switching on the light or that I

⁴ as it is in Savage's *Foundations of Statistics*, *ibid.*

⁵ cf. G.E.M. Anscombe, *Intention*, Oxford 1957.

prevent the burglar to commit a crime. In many cases, it is enough to know which effects my actions have on the state of the world, but in deontologically relevant situations this is not enough. Some action of mine a may with non-zero probability result in an action b which means that I break my promise. If I am a deontologist I may think that breaking promises is not right, even if it has good consequences regarding the state of the world. Many persons, myself included, have such a normative belief, and we should not exclude that some of these persons are nevertheless rational. But then it is relevant for such a person to know the probability of her breaking the promise as a result of doing b by doing a .

If I accomplish an action a and this action will result in action b or action c and if a is of no intrinsic value, I should be indifferent between a and the respective probability distribution between b and c which a brings about. The a -conditioned probabilities of b and c should determine my decision. This remains true also if b and c are comprehensively described and valued under deontological criteria. The probabilistic axioms of decision theory describe how the a -conditioned probabilities of b and c should determine my decision. A rational person has coherent preferences regarding probabilistic relations between actions.

If I prefer action u to v and by doing a I accomplish either u or v and by doing b I also accomplish either u or v and if a and b are deontologically neutral in the sense explained in the last paragraph, and if u is more probable if I do a than if I do b , then it would be irrational to do b . To do a in such cases is a minimal requirement of coherence that deontologically motivated rational preferences should fulfil. Let us call this assumption the *Extended Monotonicity Axiom*.

If I prefer action u to v and v to w for deontological reasons, then there is a unique probability distribution on u and w such that I am indifferent between this prospect and v . This prospect can well be caused by an action although the consequences are again actions. This is another minimal requirement of coherence that deontologically motivated rational preferences should satisfy. Let us call this assumption the *Extended Continuity Axiom*.

If we add the assumption that the preferences of a deontologically motivated and rational person are *complete* (*Extended Completeness Axiom*) and *invariant* according to the *Probability Calculus* (*Extended Substitution Axiom*), we can conclude by simple mathematical deduction that her preferences can be represented by a subjective value function⁶. A *rational person following deontological reasons optimizes subjective value*⁷, because her preferences adequately described are Ramsey-coherent⁸.

IV Remarks on Interpretation

This result needs some interpretation with which I want to end this article.

(1) It goes without saying that our result does not show that consequentialism is indispensable. John Broome, however, used an argument which has some similarity to

⁶ It is unnecessary to demonstrate this mathematical deduction here, because our requirements of minimal coherence regarding deontologically motivated preferences are equivalent with the axioms of the utility theorem, see John von Neumann/Oskar Morgenstern, *Theory of Games and Economic Behavior*, Princeton 1947 and R. Duncan Luce/Howard Raiffa, *Games and Decisions*, New York/London/Sydney 1957, chap 2.

⁷ *Optimizes subjective value = maximizes expected subjective value.*

⁸ Since Frank P. Ramsey was the first who developed the idea, almost unnoticed, of the utility-theorem, I call the property of a preference-relation, which consists in meeting all the decision-theoretic axioms "Ramsey-Coherence". This terminology has the additional advantage that the emphasis is laid on coherence, which is what the axioms are about.

mine in order to show that consequentialism is conceptually unavoidable⁹. If the reasoning above is correct, this is a *non-sequitur*. The mistake is due to a confusion of *optimizing subjective value* and *optimizing consequences*. If consequences are states of the world (or sets of states of the world, or propositions referring to states of the world) then optimizing subjective value is compatible with *not* optimizing consequences. Following deontological rules is compatible with optimising subjective value, even more: following deontological rules *rationally* requires optimising subjective value.

(2) The rational deontologist has coherent preferences for actions and states of the world alike. The deontologist differs from the consequentialist in that her preferences in general cannot be *reduced* to her preferences between states of the world. The rational deontologist is not hostile to consequentialist reasons, in many cases there are no deontological reasons which have to be considered in order to take a good decision. These cases allow to extend the deontologist's subjective value function to states of the world.

(3) Disentangling the value of consequences (states of the world) and the value of actions is necessary in order to incorporate deontological reasons in the theory of rationality. At the same time it allows for solving some seeming paradoxes of rational choice theory, e.g. the Prisoner's Dilemma. The moral requirement to act cooperatively can now be fulfilled without falling short of optimizing subjective value. A comprehensive description of strategies includes the structural traits of the game-format in which the strategy is embedded. Choosing the cooperative strategy in a Prisoner's Dilemma situation does not *ipso facto* violate Ramsey-coherence any more.

⁹ *Weighing Goods. Equality, Uncertainty, and Time*, Oxford 1991, chap. 5.

(4) Regarding the class of deontologically motivated actions there is still a logical space for weighing processes of different kinds. A fully rational person chooses her actions such that she optimizes the state of the world within the constraints given by deontological rules she accepts. In some cases the disadvantages regarding the causal consequences for the state of the world may be so severe, however, that violations of deontological rules are justified. Deontological rules themselves have to be weighed against each other and probabilities of actions which result from my actions have to be taken into consideration, too.

(5) Subjective value in this extended sense represents the whole of a rational person's reasoning process, including deontological reasons. *Extended subjective value* requires a *comprehensive description* of actions that allow for weighing consequentialist and deontological reasons alike. Extended subjective value corresponds adequately with the complexities of practical reasoning, whereas standard decision theory falls short of these complexities. Our teleological intuition is preserved insofar as rational action, including rational deontological action, optimizes subjective value.

(6) An action is rational if it is motivated by adequate reasons, and if it is adequately related to these reasons. The rationality of action cannot be judged in isolation. Practical rationality is linked to the wider context of the acting person's beliefs and preferences. We expect that these beliefs and preferences meet some requirements of coherence. The extended axioms of decision theory are criteria for these requirements. A substantial theory of rationality is a theory about good reasons to act. Decision theory is a formal theory of rationality which defines some minimal, but indispensable requirements of coherence. If these requirements are met, beliefs and preferences can be represented by two real-valued functions of subjective probability and subjective value. This numerical

representation should not be confused with substantial rationality. Rational deontological action optimizes subjective value in the extended sense, that has been introduced in this article. The deontologist optimizes subjective value if her preferences are coherent. Since preferential coherence is a necessary, but not sufficient, requirement of rationality, we know that the rational deontologist optimizes subjective value, but we do not know whether the deontologist who optimizes subjective value is rational.

(6) If my argument is valid, decision theory is an all-embracing, but not a full-blown theory of practical rationality. Decision theory is *all-embracing* insofar as all rational actions comply with decision-theoretic criteria. It is, however, not a *full-blown* theory of practical rationality, since decision theory deals only with requirements of preferential coherence. Preferential coherence as it is defined by decision theory is a precondition of substantial rationality in general (deontological reasons included), it is not substantial rationality itself. A good theory of practical rationality should be holistic and coherentist: A good theory of practical rationality should not exclude any essential element of the reasoning process from the start like deontological reasons (it should be holistic) and it should allow for a reasoning process covering all elements (it should be coherentist), no element should in principle be excluded from being weighed against other elements. Decision theory can indeed meet these requirements if it is based on *extended preferences* and *extended axioms of coherence* referring to *comprehensive descriptions of actions*.

The author is Professor for Political Theory and Philosophy at Munich University.

Julian.Nida-Ruemelin@lrz.uni-muenchen.de