Ethic Theory Moral Prac (2010) 13:485–496
DOI 10.1007/s10677-010-9228-9

Introduction: The Relevance of Rational Decision Theory for Ethics

Christoph Lumer

Accepted: 21 April 2010 / Published online: 28 May 2010

1 The Many Uses of Rational Decision Theory in Ethics

Rational decision theory (including rational game theory[1]) has played important roles in more formally oriented ethics, from providing the definition of 'utility' in welfare ethics to much more substantial contributions, like the justification of morals in game-theoretic reconstructions of contractarian ethics. Further uses of rational decision theory in ethics include: the justification of morals by means of rational decisions in an original position; the axiomatic justification of utilitarianism; risk ethics etc. So there are many ways in which rational decision theory has actually been used in ethics. These various endeavours are rather intricate, and it is far from clear which role(s) rational decision theory *should* play in ethics. There is a role for *empirical* (psychological or economic) decision theory in ethics too. However this introduction as well as this whole issue of *Ethical Theory and Moral Practice* deals with *rational* decision theory only; therefore, the addendum "rational" in the following will often be omitted. During the last decades, several proposals have been made to justify certain systems of morals in a decision-theoretic fashion (e.g. by Harsanyi, Rawls, Gauthier, Broome); and, because of the purely formal character of rational decision theory, such proposals are held to be rather strong justifications of morals. However, as is usual in ethics, all these proposals have been contested—in part precisely because of their too formal nature.

How much rational decision theory can and should contribute to ethics largely depends on which kind of justification of morals (or parts of morals) by decision-theoretic means in the end turns out to be valid and sound. The various justificatory uses of decision theory in ethics can be ordered according to the strength of the claims these arguments try to prove. The stronger the most ambitious justificatory project that in the end turns out to be valid and sound, the more important the role of rational

[1] Rational game theory here is conceived as a branch of rational decision theory, namely that branch that deals with decisions where the consequences of one's options depend (at least partially) on other rational players' decisions.

C. Lumer (✉)
Università degli Studi di Siena, Dipartimento di Filosofia, Via Roma, 47, 53100 Siena, Italy
e-mail: lumer@unisi.it

decision theory in ethics should be. The strength of decision-theoretic justificatory claims may be defined (at least) two-dimensionally, in terms of: (i) how much of morality can be justified by means of decision theory—the total enterprise of morality, including specific moral values, moral norms, institutions, virtues etc., or only special parts of it, like certain values or dealing with risk—, and (ii) how exclusive the role of decision theory is in this justification, i.e. can (certain) morals be justified by relying on the formal precepts of decision theory and persons' general (coherent) preferences (whatever they are) only or have further premises or sources to be used, like certain altruistic features of one's preferences or some general kind of moral intuition? The more of morality that can be justified by decision theory and the fewer further sources are used, the stronger is the justificatory role of decision theory.

According to these ideas about the strength of the justificatory role of decision theory (and intuitively dealing with conflicts between the two dimensions of strength), the most important decision-theoretic endeavours to justify morals can be brought in the following order, beginning with the strongest justification of morals:

1. *Cooperation ethics:* The best-known, strongest and meanwhile traditional approach aims at a game-theoretic, contractualist justification of an ethic of cooperation on the basis of rational interests in the outcomes of cooperation (e.g. Gauthier 1986; Skyrms 1996, 2004; Binmore 1994; 1998, 2005).

2. *Ethics of rational moral value:* An intuitively more moral approach starts with the definition of moral value functions as those altruistic parts (aspects, attributes) of our rational utility functions that are based e.g. on moral sentiments (like sympathy or respect); in a full-blown version of this approach, the definition of moral value is complemented by the (construction and) justification of norms, institutions etc. that realize moral value and make moral action rational (e.g. Rescher 1975; Margolis 1982; Lumer <2000>/2009, 2011).

3. *Axiomatic justification of utilitarianism:* In a quite formal approach, the shape (in particular the utilitarian shape) of welfare ethics has been justified on the basis of the axioms of rational decision theory (and of the aim to want to have a welfare ethics) (Harsanyi <1955>/1976: 7; 10–13; 1977a: 48–83; 1977b; Broome 1991).

4. *Ethics of impartial rational decision:* Alternatively, morals have been justified via an impartial rational decision, where impartiality is operationalized by uncertainty about one's role, e.g. in an original position with a veil of ignorance (Harsanyi <1955>/1976: in particular 13 f.; Rawls 1971).

5. *Risk ethics:* Risk ethics can be developed, relying on the handling of risk in rational decision theory for risky choices (Fritzsche 1986; other approaches to risk ethics; Hansson 2003; Lewens 2007).

These five approaches are attempts to justify certain kinds of morals (or only parts of them). The use of rational decision theory in ethics goes beyond them and includes, in a more subordinate role, its application for systemizing and defining parts of ethics:

6. *Definition of 'moral value':* The technical instruments of rational decision theory, in particular subjective expected utility theory, may be used to define '(expected etc.) moral value', even independently from its decision-theoretic justification.

7. *Measuring personal utility:* Utility theory as a branch of rational decision theory can provide the quantitative concept of '(personal) utility' and the means to measure utility as specifications of the concept of 'personal good' presupposed in welfare ethical definitions of moral value.

These uses of rational decision theory in ethics shall now be explained and discussed.

2 Cooperation Ethics—Strengths and Problems

Game-theoretic ethics of cooperation are the strongest endeavours to justify morals by rational decision theory because they try to justify (i) a complete moral (ii) with the weakest premises: the precepts of game theory and—rather arbitrary and self-interested—personal preferences. One of the central ideas behind this economical approach is a certain thesis, which is generally important for decision-theoretic approaches to ethics: namely the *rationality requirement of morals:* The demands of an adequate morality must be such that it is decision-theoretically rational to fulfil them. An argument for this rationality requirement is this. First, for being adequate, a morality must be motivationally effective (i.e. lead people (who are informed about it as well as about its justification) to moral behaviour) and epistemically rational (which implies that its justification may not rely on false or insufficient information). Second, the motivational effectiveness of instructions or recommendations, at least in the long run, requires them to be based on the agent's preferences and, because of epistemic rationality, on all his considered preferences.[2] Third, rational decision theory just defines 'rationality' in this way. Finally, fourth, the demands of practical rationality thus defined, from the agent's point of view, "beat" those of morality: the rational requirements are more inclusive, they include *all* of the agent's concerns; morality is just one of them.

The straightforward way, followed by cooperation ethics, to fulfil the rationality requirement of morals is to construct moral demands as particular demands of (decision-theoretic) rationality. However, to avoid leading directly to moral egoism, some fairly general internal or external restriction, or fairly common and characteristic field of application, has to be found that brings rationality more in line with common morality. Now, this special but rather wide field of application has been identified with situations of risky cooperation. In such situations, there is room for morality because the subject does something good to his partner (minimal altruism); and there is room for rationality because cooperation requires that the partner does something good for the subject—which outweighs the costs of the altruistic part—; finally, all this is not immediately egoistic but requires some moral device because the cooperation is risky and tempting at the same time: the other may cheat me in not providing her cooperative share, whereas I do; and I may cheat my partner by doing the same with her, which I should not do from a moral viewpoint. The main task of cooperation ethics is to find ways to bring morality and rationality in line in this challenging situation. One such challenging situation is the prisoner's dilemma, where the moral aim is mutual cooperation. However, there are many other types of games with possible but problematic cooperation for which cooperation ethics tries to develop some general game-theoretic solution;[3] Gauthier's proposal of minimax relative concessions together with restricted maximizing is such a solution (Gauthier 1986: chs. V–VI).

The three main approaches in cooperation ethics to bring moral cooperation requirements and rational requirements of maximizing one's utility in line rely on: (i) sanctions by a third party for non-cooperation, (ii) iteration of cooperation situations with the same partner (if I cheat you in a one-shot game, where later on I will never see you again, there is no (rational) incentive not to cheat you; however, this situation changes if the game is repeated because, having cheated you, you (most likely) will not cooperate with me in the next round; so by cheating I lose possibilities of further cooperation); and, finally, (iii) refusal of cooperation by other people: other people may learn about my cheating and

[2] These two premises, of course, are quite substantial and should (and can) be justified. However, here is not the space to defend them (but cf. Narveson's and McClennen's contributions to this special issue, or e.g. Lumer <2000>/2009: 30–42). Anyway, most sympathizers of a decision theoretical approach accept them.

[3] Cf. e.g. Narveson's contribution to this special issue.

hence untrustworthiness, therefore, in the future they too may not be willing to cooperate with me—not necessarily for punishing me but in any case as a measure of prudence (‘this guy will probably cheat me, so better not to cooperate with him’)—so that I will lose further possibilities of cooperation. Blends of these three ways are possible: the cheated partner (of (ii)) as well as third parties (iii) may consider their refusals to cooperate with me also as sanctions for keeping up a certain moral order (i).

However, these strategies of cooperation and mechanisms of enforcing them are still rather fragile and may not work in many instances: it may be rather sure that we will not enter into a further situation of possible cooperation or that no other person will know about my cheating; the next situation of possible cooperation might be quite different; you may be depending too much on my cooperation and therefore swallow my occasional or frequent non-cooperation; a hierarchy of external sanctions finishes somewhere etc. In addition, the strategies of cooperation may be costly: there may be an arms race to secure to have the best threat-point or starting position in cases of non-cooperation; punishing may be directly costly, or risk being replied by revenge; it may even lead to social isolation (not everywhere in the world do people like and support defenders of moral correctness) etc. Finally, the strategies of cooperation just expounded seem not to include and explain several mechanisms actually present and important in cooperation: no small number of people cooperate even when they do not see any chance of reciprocation, or they punish non-cooperation at personal costs.

These weaknesses of traditional game-theoretic strategies of cooperation have stimulated some defenders of cooperation ethics to introduce unorthodox amendments or alterations, e.g. virtues or fairness considerations, into cooperation ethics. (The contributions of McClennen and Verbeek to this special issue are of this sort.)

Apart from these internal criticisms and advancements, there are external criticisms, which aim at refuting cooperation ethics altogether or at least at assigning it only a minor role in ethics. From the viewpoint of substantive ethics, ethics of cooperation are rather weak (in the sense of demanding little): they are only minimal or business ethics (applicable even within the Mafia), disregarding fairness and distributive justice. Beings unable to cooperate are not included among their direct beneficiaries: small children, the very old, strongly handicapped, future generations, animals. Those able to cooperate receive advantages according to the extent that they can and will provide advantages to others; so the strong, talented, wealthy and mighty receive a lot; the weak, untalented, poor and powerless receive little and it costs them much. Cooperations are rational only if they are Pareto improvements, i.e. better for all partners; this excludes redistribution. From a structural viewpoint, cooperation ethics lack some important features of common ethics: Cooperation ethics provide a *norm* of cooperation but they do not provide any supra-personal *moral value function* and moral valuation. They cannot say e.g. that a certain cooperation was rational for everybody involved but nonetheless unfair; they just stick with personal valuations. And because moral valuation is the origin of moral feelings in the strict sense (like indignation, resentment, shame or guilt) cooperation ethics have no place for moral feelings. Furthermore, because these feelings have motivational value, cooperation ethics disregard important motives for acting morally. Cooperation ethics, until today, are bound to game theory, which provides rational *strategies of action* in situations of interaction. This prevents them from considering other devices of morality: social norms (in the usual sense of the word), institutions, constitutions, economic orders, attitudes and faculties other than dispositions to act (e.g. sensibility and empathy, (moral) wisdom).

Cooperation ethicists reply to these criticisms that they beg the question. They presuppose moral standards, devices etc. that have not been shown to be rationally justifiable. However, even this correct reply does not imply that these standards, devices etc. are *not* rationally justifiable. Whether they are, i.e. whether a rational justification of a stronger kind of ethics is possible, so far remains an open question for further research and discussion.

3 The Project of an Ethic of Rational Moral Value

Ethics of rational moral value, unlike cooperation ethics, take a moral value function to be central to morality and, therefore, systematically start by determining this value function and then constructing the rest of ethical theory on this basis. This moral value function cannot be identical with one's total (personal) utility function; however, to be rational and at least minimally motivating, it must be identical with some part or aspect or attribute of one's total rational utility function. The first question, then, is what does this moral aspect consist in? The most important answers to this question are: it is the altruistic aspect and/or the interpersonally shared aspect of one's general utility function—where "shared" shall mean that the same (and not only analogous) objects are valued more or less identically. Some reasons for this answer are e.g. intuitionistic ('this is usually considered as the moral part of our utility functions') or functionalistic ('thus we have a common value function, with the help of which we can decide about common projects') or psychological ('thus we have defined a self-transcendent viewpoint, which helps persons to decentre, to find sense, long-term satisfaction and peace of mind'). The most important components of our utility function that have the required features are the aspects stemming from sympathy (including its negative parts: compassion and pity) and from respect for other beings (persons, animals, nature and even cultural objects) (Lumer 2002). If e.g. two moral subjects, s_1 and s_2, see another one, h, suffering or in a plight, they may both feel compassion and as a consequence prefer and be motivated (somewhat) to improve h's situation; in addition, anticipating h's possible suffering and their own compassion, they may prefer and be motivated (somewhat) to prevent this suffering. The technique for quantitatively determining the appertaining aspect of desirability functions—which then define the moral value function—is provided by rational decision theory: it is the multi-attribute utility theory. Justifications of morals provided by ethics of rational moral value assign decision theory a weaker role in justification than ethics of cooperation because the former essentially rely on a strong empirical premise, namely the existence and particular shape of an adequate aspect of our total desirability function.

Moral value is only the first half of morality, which has to be complemented by a theory of moral norms, institutions, virtues etc. The method to proceed in this second part of the ethic should be straightforward. Moral norms, institutions, virtues etc. are instruments for realizing moral value; accordingly, they should be constructed in an instrumentalist fashion, namely in such a way as to realize much or maximum moral value. The justificatory crucial task in this second part of ethics of rational moral value is again decision-theoretic: to determine the morally best of these instruments.

However, ethics of rational moral value have to fulfil the rationality requirements of morals too (cf. compare above, Section 2)—which gives rational decision theory a further role in this approach. Therefore, the moral instruments to be constructed have to be designed in such a way that it is decision-theoretically rational to follow the moral demands implied by them. This is a further condition that these instruments have to fulfil beyond the condition to realize as much moral value as possible. To understand its importance, consider the case of moral norms and obligations. The maximizing condition alone would simply lead to the (optimific) obligation always to do the morally best. Such a simplistic obligation, however, would not fulfil the rationality requirement of morals because moral value is only one aspect of the subject's total desirability and is often not decisive. Hence, moral norms and obligations have to be designed in a much more complicated way, in particular by including sanctions for norm violations (which for the individual provide further reasons to act morally) and—which is much more difficult—by including ways that may lead to a social adoption of the norm that entails a social practice of administering such sanctions.

The ethic of rational moral value has found much less attention and fewer supporters than ethics of cooperation. As a consequence, that ethic is not elaborated so much: a big part of it is more a project than a theory. The real problem of this approach is whether it succeeds in fulfilling the rationality requirement of morals.

4 Axiomatic Justification of Utilitarianism and Ethics of Impartial Rational Decision—Moral Intuitions Plus Decision Theory

In ethics of rational moral value, for justifying morals, further premises had to be added to the decision-theoretic rules, namely empirical premises about the existence and shape of common altruistic motives and the resulting aspect preferences. In the axiomatic justification of utilitarianism and in ethics of impartial rational decision, again further premises must be added to the decision-theoretic rules, this time, however, certain *moral intuitions*.

The most important moral intuitions used in Harsanyi's axiomatic justification of utilitarianism are: i. an adequate ethic has to be welfaristic, i.e. an ethic where 'moral value' is defined as an aggregation of the beneficiaries' utilities; ii. the personal utility functions and the moral value function, both, shall fulfil the standard axioms of subjective expected utility theory and they shall fulfil them combined (Harsanyi <1955>/1976, 10–13). The trick of the argument then consists in exploiting the latter feature: In standard subjective expected utility theory, probabilistic outcomes are weighted linearly; certain outcomes with a given (total) utility are regarded as being equally good as and interchangeable with lotteries with the same expected utility; in addition, these requirements have to hold for the personal utility as well as for the moral value of this personal utility. All these requirements can be fulfilled together only if the moral value is the sum of the moral values of the personal utilities and if the moral value of a personal utility is a linear function of, or simply identical to, this personal utility—which is the utilitarian value function.[4]

[4] The second consequence can be shown rather easily. Let $V_i(x)$ be the *i*-th component of the moral value function, which represents the moral value of person *i* obtaining the (personal) utility or expected utility (U_i) x for some event or lottery; as substitutions for "*x*" we allow the cardinal utility values themselves or the events (or lotteries) that produce these cardinal utilities. Let k be a number between 0 and 1 (i.e. $0 \leq k \leq 1$); let u_i be *i*'s utility for some event; and let $\langle p;x;(1\text{-}p);y \rangle$ be the lottery where one obtains x with probability p, and otherwise y. What has to be shown to prove the linearity claim then is this: $V_i(u_i \cdot k) = k \cdot V_i(u_i)$: if u_i is diminished by the factor k, the respective moral value has to be diminished by the factor k too. Now, according to the van Neumann/Morgenstern axioms, $u_i \cdot k = U_i \langle k;u_i;(1\text{-}k);0 \rangle$. Because these two things are equally good for the subject, according to Harsanyi's assumptions (which are, however, contested), they should be equally good also from the moral point of view, hence $V_i(u_i \cdot k) = V_i \langle k;u_i;(1\text{-}k);0 \rangle$. And because even from the moral point of view, probabilistic outcomes should be weighted linearly with their probability, the moral value of the latter lottery should be identical to its expected moral value, hence: $V_i \langle k;u_i;(1\text{-}k);0 \rangle = k \cdot V_i(u_i) + (1\text{-}k) \cdot V_i(0) = k \cdot V_i(u_i)$, q.e.d. (Cf. Harsanyi <1955>/1976, 10–12.)—This axiomatic proof of utilitarianism has been criticized by attacking its axioms. Even a welfarist, who accepts the welfaristic setting, and hence the first part of the axioms, can refuse the combined application of all van Neumann/Morgenstern axioms to the personal utility function as well as to the moral value function. This means that he can in particular state that risky prospects have to be dealt with either on the personal or, preferably, on the moral level, but not on both levels at the same time. The second possibility, i.e. moral treatment of risky prospects only, e.g. implies that from $u_i \cdot k = U_i \langle k;u_i;(1\text{-}k);0 \rangle$ no longer follows $V_i(u_i \cdot k) = V_i \langle k;u_i;(1\text{-}k);0 \rangle$.

The most significant moral intuition used in ethics of impartial rational decision is impartiality, which is operationalized as complete ignorance about one's personal characteristics: one can just be identical with any subject in this society and one does not know anything about with whom. Moral principles then shall be chosen under these restrictions of impartiality. To have to choose between options with possible consequences to which one cannot even assign probabilities, in decision-theoretic terms is a decision under uncertainty so that, in the impartiality setting, maxims for decisions under uncertainty should be applied. Now, there are several competing proposals for such maxims. Rawls has based his proposal on the maximin principle (choose the option whose worst possible outcome is better than the worst possible outcomes of the other options), thus reaching his difference principle (Rawls 1971). Harsanyi, criticizing maximin as too pessimistic and overweighting, without any reason, the possibilities with the worst outcomes, has based his proposal on the principle of sufficient reason, which requires treating all the possibilities as, fictiously, equi-probable (of course, only in situations of uncertainty), thus reaching utilitarianism (the utilities of each of *n* persons have to be weighted by $1/n$, i.e. the fictious equi-probability) (Harsanyi <1955>/1976: 13 f.). Still others followed Harsanyi in assigning equal probabilities to the various possibilities; however, they denied that risky prospects (from a personal point of view) should be simply weighted according to their probabilities. Instead of this, they proposed, in a risk-aversive spirit, to give some more, however not infinitely more, weight to avoiding negative outcomes and less weight to improving one's lot. (This can be done by assigning to the utilities of equi-probable outcomes risk-weighted values by means of a monotonously increasing but concave weighting function, i.e. a function where less and less additional risk-weighted value is assigned to additional utilities, the higher the utility level already reached.) This kind of risk weighting together with the impartiality setting leads to a prioritarian welfare function, where improving persons' lot matters more, the worse off these people are.[5]

The axiomatic justification of utilitarianism as well as ethics of impartial rational decisions essentially uses certain moral intuitions as premises. This makes them weak justifications because an addressee of this justification for accepting it must have accepted these premises; and for to be a rational justification even for those addressees who have already accepted these premises, this acceptance has to be justified. In all other cases, such an intuitionistic justification is begging the question. This problem of an intuitionistic justification does not only regard the special content of these moral premises (e.g. morality as an impartial view on the world), it also regards the very undertaking of justifying morals and acting morally in a motivating way as a whole. However, this is a problematic feature of any intuitionistic justification of morals. A more specific problem of the two discussed intuitionistic approaches to justify morals with the help of additional premises taken from rational decision theory is that some more or less technical precepts (or at any rate precepts that have nothing to do with morals) have a crucial role in determining the shape of the resulting moral. In Harsanyi's axiomatic justification of utilitarianism, the axiomatic arrangement to permit the combined application of the van Neumann/Morgenstern axioms to individual utilities as well as to social values plays the decisive role. In ethics of impartial rational

[5] A justification of the prioritarian idea by means of an impartial rational decision has been provided by: Atkinson and Stiglitz 1980: 340; Hurley 1989: 368–382. The name "prioritarian view" is younger, it goes back to: Parfit 1995. However, prioritarianism can also be justified quite differently, namely in an ethics of moral value as the moral value function stemming from sympathy: cf. Lumer <2000>/2009: ch. 7; 2011.

decision, on the other hand, the assumptions about the precepts for decisions under uncertainty are decisive for the choice between the difference principle, utilitarianism and prioritarianism. One may hold instead that such important moral questions should not be determined on the basis of such completely extraneous considerations. Whether I e.g. want to be a follower of Rawls, a utilitarian or a prioritarian does and should not depend on my risk behaviour; after all, there certainly exist strongly risk-aversive utilitarians and risk-neutral prioritarians and Rawlsians whose attitudes seem to be completely coherent.

5 Risk Ethics

Rational decision theory discusses (mostly) decisions under risk, i.e. decisions where the consequences of options (at least in part) are not certain, but where the decider can assign a probability to these consequences. This concept of 'risk' only partially overlaps with the respective concept in risk ethics (and in ordinary language); the latter speaks of *negative* consequences only and it includes decisions where we have even less than probabilistic information (no probabilities, no clear idea of the possible consequences). However, some of the latter cases may be considered as decisions under uncertainty, which again are discussed in decision theory, so that the objects dealt with in these two disciplines are sufficiently overlapping. As a consequence, decision theory, at least *prima facie*, should have something to contribute to risk ethics.

What it can contribute are, of course, some proposals that could at least *inspire* risk ethics: i. its orthodox suggestions about individual decisions under risk, namely to maximize expected utility; ii. the suggestions for and discussion about heterodox, i.e. non-linear, weighting of probabilities in decisions under risk with extreme probabilities or extreme utilities; iii. the suggestions for and discussion about how to decide in situations under uncertainty. However, it would be fallacious simply to transfer the decision-theoretic proposals to risk ethics. This is so because rational decision theory deals with individual choices and weighing risks (in the ordinary sense) against chances for one and the same person, whereas risk ethics discusses individual as well as social choices and weighs the risks and chances of some people against (different) risks and chances of other people. So, there is a further dimension where weighing can take place. To put it another way: risk ethics, unlike decision theory, have to consider the social distribution of risk too. Simply transferring the decision-theoretic solutions of risky choices to risk ethics—as e.g. Harsanyi does in his axiomatic proof of utilitarianism—ignores, deliberately or naively, the distribution problem. Of course, one can determine, as utilitarians do, that distribution does not matter; however, this would not be a consequence of decision theory but a separate moral decision and not justifiable by the rules of rational decision theory.[6] So, the role of rational decision theory in the *justification* of risk ethics, in contrast to first appearance, seems to be rather limited.

Utilitarianism, as just said, neglects distribution of risk and has been criticized for this, e.g. as disregarding the separateness of persons. Common moral intuitions, on the other hand, give great weight to questions of distribution in a way that is not easily (if at all) to be formalized in welfaristic terms. Think e.g. of the usual norm that, to allow the use of technologies with a certain (physical, biological etc.) impact, some residual risk of dying may be imposed on non-usufructuaries of this technology but this risk has to remain under 10^{-6}/year. First, this norm distinguishes between users, usufructuaries and non-usufructuaries of the technology, which is usually disregarded in welfare ethics.

[6] For further criticisms, see also Hansson's article in this special issue.

Second, the norm allows, technically speaking, options with the possible outcome $\langle u_1;...;0;...u_n\rangle$ (where 0 is the utility of death), but the probabilities of such outcomes have to be under P=10^{-6}/year; prospects not fulfilling this feature are not valued rather low, they are simply forbidden. It is hard to model these features by a welfare function. However, there may be a way out here for a welfarist approach—apart from simply refusing such intuitive ideas as not rationally justifiable—along the following strategy: this kind of risk ethical intuitions does not refer to the moral *value* of such prospects but to moral *norms* of what to do with prospects of a certain kind; such intuitions therefore have not to be modelled in the axiological but in the normative part of the theory. Of course, this strategy presupposes that the normative part of ethics does not just consist of the (optimific) prescription always to maximize moral value but allows much more complicated ways to refer to moral value. According to such a strategy, the intuition about residual risk would be interpreted as a special norm that guarantees to everybody some sort of minimum standard with respect to the "technology dependent" probability of survival of at least 1-10^{-6}/year; this guarantee may not be violated even to reach higher moral values.[7] Such a norm may be in line with some other norms guaranteeing basic rights. If, according to this strategy, many of the risk ethical questions have to be dealt with in the normative and not in the axiological part of ethics, then it seems very unlikely that rational decision theory, mainly thematizing individual axiology, can contribute something to the solution of these special questions.

6 Defining 'Moral Value' and Measuring Personal Utility

There is no straightforward way leading from decision theory to the basic, unprobabilized definition of 'moral value'. This definition plainly poses a new question, i.e. how to aggregate personal utilities to moral value; and the range of proposals made to this effect (utilitarianism, egalitarianism, prioritarianism, to name only a few) as well as the justifications given for them show that completely new considerations are decisive in this debate. However, this does not exclude that once these 'basic moral values' are defined, they may be subjected to decision-theoretic treatment for defining 'expected moral value' etc., which then could be used to determine the moral value of prospects with outcomes that are not certain and to decide between them in situations that do not fall under the special restrictions of risk ethics. This would require that moral values thus defined fulfil the (or some of the various sets of) axioms of rational decision theory. Such a proceeding would not imply that a welfare distribution a=(0.5;0.5) (i.e. person 1 has a utility of 0.5 as well as person 2) has the same moral value as the following lottery over welfare distributions b= $\langle$0.5; (0;1); 0.5; (1;0)$\rangle$ (this is the lottery to obtain with equal chances the welfare distribution (0;1) or (1;0)). This holds because the moral value of (0.5;0.5) need not be identical to the moral value of (0;1) (or (1;0)); only utilitarianism assumes this. An analogous case in individual decision theory is: the utilities of receiving 100,000 euros for sure or entering a lottery where one has a 50% chance of winning 200,000 euros or nothing, for most people are not identical, already because the utility of receiving 200,000 euros is not twice as high as that of receiving 100,000 euros; utility functions over income are usually concave. However, to subject moral values to decision-theoretic treatment implies that the welfare distribution c=(0;1) (or (1;0)) has the same moral value as the lottery d=

[7] Interpreted in this way, risk ethical questions are special cases for the discussion about the limits of morality (cf. e.g. Scheffler 1982; Kagan 1989).

⟨0.5; (0;1); 0.5; (1;0)⟩, i.e. a lottery where the question of who will be the winner (utility 1) and who will be the loser (utility 0) is decided by chance. However, many people think that it is fairer to assign indivisible and unmerited goods by chance than to assign them directly to somebody. So, the lottery *d*, being fairer, should have a higher moral value than the direct assignment *c*. More generally speaking, a decision-theoretic treatment of (expected) moral values may be at odds with the special moral value of chance; so it may be particularly at odds with ethics of equal opportunities.

Some possible answers to this challenge are these. First, one could abandon the project to subject moral value to decision-theoretic treatment; in particular, one might refuse to apply some axioms of decision theory to moral values because, on the social level, they acquire a completely different meaning. The moral counterpart of the independence axiom e.g. could be discarded because it does not consider the social distribution of risk. The axioms of decision theory to a great deal serve only for measuring utility. Because ‘basic moral value’ is well defined with the help of other well-defined notions, such that the basic moral value of outcomes can be determined analytically and not relying on moral preferences, there is no metrological need to have these axioms fulfilled by moral values. So, the loss by giving up the strategy may not be great. Second, one could try to integrate ideas about fairness by chance, by assigning additional moral value to outcomes of fair procedures. Dreier suggests this kind of solution (Dreier 2004: 172–175). Third, one could reject the moral intuitions about equal opportunities and stick to the project of decision-theoretic treatment of moral values. After all, one might say that what should be morally relevant are the final outcomes and not some random procedure to arrive at them. Egalitarians of well-being have e.g. objected to equality of opportunities that equality of opportunity “instead of reducing the huge gap between, say, physicians and ditch diggers, it might merely change the demographic composition of those groups” (Temkin 1993: 85 fn.). In a certain sense, positive moral value created by inserting random procedures before the social assignment of final outcomes could be bogus value. If highjackers of an aeroplane who want to underline their demands by executing one of their hostages say: “Okay, we want to be fair, let’s draw who will be shot” and do so, this would probably not add a minimum of positive moral value to their deed.

A sure use of rational decision theory in ethics, finally, seems to be the determination of personal utilities, which are needed in welfare ethics (or less formal ethics that, nonetheless, see improving persons’ well-being as the central task of morality). Sometimes, it is held that the personal good to be considered in ethics should not be a decision-theoretic utility because this utility is defined via the persons’ (uncriticized) preferences which, however, may be immoral or irrational or presumptuous (in the way that they include ideas about how other persons have to be or behave). This concern about persons’ preferences is fully justified; however, from this it does not follow that the personal good to be considered in ethics should not be decision-theoretically defined. Subjective expected utility theory in a certain sense is only a coherence theory of practical rationality and preferences, it does not judge the adequacy of these preferences. Therefore, the decision-theoretic definition of ‘utility’ can be complemented by a theory about how to select among the persons’ preferences on the basis of which utilities shall be determined; thus, one would proceed from ‘uncritical utilities’ to filtered, ‘critical utilities’. Approaches to such a selection include restrictions: to fully informed preferences (that supervive cognitive psychotherapy) (Brandt 1979: part I), to basic preferences (Lumer 1998; <2000>/2009: chs. 4–5) and to morally laundered preferences (Goodin 1986). Of course, welfare ethics should only use some form of critical utility.

7 The Contributions to this Special Issue

The articles of this special issue of *Ethical Theory and Moral Practice* deal with various of the several ways just exposed in which rational decision theory could contribute to ethics.

Jan Narveson's article "The Relevance of Rational Decision Theory for Ethics", for one thing, defends the very approach of cooperation ethics. Its idea is to (re-)construct demands of morality as requirements of game-theoretic rationality—which is more than to prove that it is rational to follow demands of morality that are, however, justified or introduced in a different way. Such a morality is based on taking into account the iteration of situations of possible cooperation, on including policing into the requirements of cooperation and on the rationality of keeping up one's reputation. For another thing, the article illustrates the applicability and ethical usefulness of this approach in several types of game even beyond prisoner's dilemma: zero-sum, battle of the sexes, pure coordination, chicken game.

Edward McClennen, in his "Rational Choice and Moral Theory", first, in a critique of several historical justifications of morals (from Socrates to Kant), tries to show that only cooperation ethics, which is based on rational self-interest, can fulfil the requirements of a *practical* justification of morals, which aims at providing reasons for decisions. Second, he analyses some problems in Nash's / Harsanyi's solution in game theory and in orthodox approaches to cooperation ethics, in particular Gauthier's, like: unreliability of iteration, costly arms race for threat potential or costly sanctions or resentment against rational but unfair agreements, which leads to boycotting such agreements. Finally, he proposes a new solution to these problems, "Full Cooperation", i.e. cooperation already during bargaining, which brings in aspects of fairness: the benefit from cooperation shall be distributed equally unless an unequal distribution would be to the advantage of all participants. The fairness of this result should lead to compliance.

Bruno Verbeek's essay "Rational Choice Virtues" tries to provide another form of amendment to orthodox cooperation ethics: a theory of the virtue of trustworthiness. On a more formal level, virtues are one usual element of morality which, in cooperation ethics, however, has so far been neglected. Therefore, the essay shall fill this lacuna a bit. On a substantial level, it is argued that previous cooperation ethics themselves need such an amendment for overcoming problems of instability of norms. In real life, these problems are resolved by means of moral emotions and virtues. The last part of the paper then elaborates a formal theory of the exact content of the virtue of trustworthiness in one-shot prisoner's dilemmata. In such situations one shall e.g. cooperate only if there are clues to the other's trustworthiness.

Christoph Lumer's article "Moral Desirability and Rational Decision" is a contribution to an ethics of rational moral value. After defending the rationality requirement of morals, cooperation ethics, which try to fulfil this requirement, are criticized, among others, as being structurally incomplete in not entailing some sort of supra-personal moral desirability. Because moral valuations are the basis for moral emotions, cooperation ethics lack this element too, thus forgoing an important source of moral motivation. In the constructive part, a bipartite ethic of rational moral value that tries to fulfil the rationality requirement of morals is sketched. First, it is shown how, by means of multi-attribute utility theory, a socially acceptable and individually motivating moral value function can be determined. Second, it is explained how, on this basis, moral norms and institutions may be socially realized.

Sven Ove Hansson's essay "The Harmful Influence of Decision Theory on Ethics" advances two criticisms against an incautious use of decision theory in ethics. First, many theorists think that social risk can simply be dealt with in a decision-theoretic fashion, i.e.

by deciding according to the expected moral value. This kind of thinking exactly ignores the moral relevance of risk, but as a consequence it has hindered the development of an independent risk ethic. Second, transferring the model for individual choices, with causal consequences of the chosen action, to the level of social decisions disregards the effects of the interplay of many actions, which cannot be attributed to any single action.

References

Atkinson AB, Stiglitz JE (1980) Lectures on public economics. McGraw-Hill, London [etc.]

Binmore K (1994) Game theory and the social contract. Vol. 1: playing fair. MIT, Cambridge (Mass.)

Binmore K (1998) Game theory and the social contract. Vol. 2: just playing. MIT, Cambridge (Mass.)

Binmore K (2005) Natural justice. Oxford U.P., Oxford

Brandt RB (1979) A theory of the good and the right. Clarendon, Oxford

Broome J (1991) Weighing goods. Equality, uncertainty and time. Blackwell, Oxford

Dreier J (2004) Decision theory and morality. In: Mele AR, Rawling P (eds) The Oxford handbook of rationality. Oxford U.P., Oxford, pp 156–181

Fritzsche AF (1986) Wie sicher leben wir? Risikobeurteilung und -bewältigung in unserer Gesellschaft. Verlag TÜV Rheinland, Köln

Gauthier D (1986) Morals by agreement. Clarendon, Oxford

Goodin R (1986) Laundering preferences. In: Elster J, Hylland A (eds) Foundations of social choice theory. Cambridge U.P., Cambridge, pp 75–101

Hansson SO (2003) Ethical criteria of risk acceptance. Erkenntnis 59:291–309

Harsanyi JC (<1955>/1976) Cardinal welfare, individualistic ethics, and interpersonal comparisons of utility. J Polit Econ 63(1955):309–321. - Reprinted in: Idem: Essays on ethics, social behaviour, and scientific explanation. Reidel, Dordrecht / Boston 1976, pp. 6–23

Harsanyi JC (1977a) Rational behavior and bargaining equilibrium in games and social situations. Cambridge U.P., Cambridge

Harsanyi JC (1977b) Morality and the theory of rational behaviour. Soc Res 44. - Reprinted in: Sen A, Williams B (eds.) Utilitarianism and beyond. Cambridge U.P., Cambridge 1982, pp. 39–62

Hurley SL (1989) Natural reasons. Personality and polity. Oxford U.P., New York

Kagan S (1989) The limits of morality. Clarendon, Oxford

Lewens T (ed.) (2007) Risk. Philosophical perspectives. Routledge, London

Lumer C (1998) Which preferences shall be the basis of rational decision? In: Fehige C, Wessels U (eds) Preferences. De Gruyter, Berlin, pp 33–56

Lumer C (<2000>/2009) Rationaler Altruismus. Eine prudentielle Theorie der Rationalität und des Altruismus. 2nd, supplemented ed. Mentis, Paderborn ^{2}2009

Lumer C (2002) Motive zu moralischem Handeln. Analyse & Kritik 24:163–188

Lumer C (2011) Prioritarian welfare functions—an elaboration and justification. Forthcoming. - Web publication: http://mora.rente.nhh.no/projects/EqualityExchange/ressurser/articles/lumer1.pdf

Margolis H (1982) Selfishness, altruism, and rationality. A theory of social choice. Cambridge U.P., Cambridge [etc.]

Parfit D (1995) Equality or priority? The Lindley Lecture, University of Kansas, November 21, 1991. University of Kansas, Kansas. - Reprinted in: Clayton M, Williams A (eds.) The ideal of equality. St. Martin’s Press 2000

Rawls JB (1971) A theory of justice. The Belknap Press of Harvard U.P., Cambridge, Mass. - New edition: A Theory of Justice. Revised edition. Oxford U.P., Oxford [etc.], 1999

Rescher N (1975) Unselfishness. The role of the vicarious affects in moral philosophy and social theory. Univ. of Pittsburgh Pr, London

Scheffler S (1982) The rejection of consequentialism. A philosophical investigation of the considerations underlying rival moral conceptions. Revised edition: Clarendon, Oxford ^{2}1994

Skyrms B (1996) Evolution of the social contract. Cambridge U.P., Cambridge

Skyrms B (2004) The stag hunt and the evolution of social structure. Cambridge U.P., Cambridge

Temkin LS (1993) Inequality. Oxford U.P., New York