

Moral Choice Without Moralism

Erich Rast

Abstract

Curiously, in ordinary rational choice and multi-criteria decision making moral attributes of alternatives play no role that sets them apart from other kinds of attributes. In this article, several ways of defining a decision as moral on the basis of attributes that have already been determined as being morally relevant are investigated that do not trivialize the distinction by always giving priority to moral attributes. Instead, a threshold view is proposed that allows the application of moral rules to be permeable or 'soft' in dependence of what is at stake in a decision situation. The approach forces a decision maker to make thresholds of moral relevance explicit and may also be useful for group decision making.

Keywords: decision making; moral vs. prudential value; permeable deontic rules; excusable actions; rational choice

A moralist is a pedant who insists on conforming to strict moral rules in each and every case. This attitude is often a symptom of 'moral tyranny', which arises '...when the claim to superior knowledge of good and evil lacks justification' (von Wright 1963: 189). Both attitudes can lead to a rejection of the underlying moral codes and undermine morality in general. How can a decision maker act morally without becoming a moralist?

In this article, I will address one aspect of this question from a purely decision-theoretic perspective by giving a tentative answer to a narrower and more specific question: What constitutes a moral decision, provided that in a given decision situation a set of morally relevant attributes can be identified and that these attributes do not automatically and unconditionally override all non-moral attributes? To answer this question in a decision-theoretic setting I propose a theory with lexicographic thresholds that allows a decision maker to deviate from the prescription made implicitly by the morally relevant attributes whenever the stakes are low, but requires the decision maker to follow the morally relevant attributes when the stakes are high. Conditions are laid out under which a decision is morally acceptable (or, short: 'moral') in this setting. The approach is based on the assumption that the decision making methodology is principally adequate for making moral decisions and is in principle compatible with much of what has been said in the literature on Practical Reasoning. Albeit controversial, this view will not be challenged here but rather supposed for the sake of this article. However, I believe that another assumption needs to be explained in more detail, namely the one that moral attributes do not unconditionally outrank non-moral attributes. In deontic logic a distinction is often made between actions that need to be done (required/obligatory actions; actions that are one's duty), actions that are permissible and actions that are forbidden. Using these distinctions as a basis, the thesis boils down to the claim that a course of action that one is permitted to do may sometimes outweigh a course of action that one ought to do if not much is at stake from a moral point of view. Another, in my point of view better way of putting this, is based on Meinong's distinction of value classes. He classifies values into four categories: meritorious (verdienstlich), correct (correct), merely excusable (zulässig), and

inexcusable (verwerflich), where the first two are good and the latter two are bad. According to this taxonomy, some courses of action may be excusable in the sense of being morally permitted despite the fact that they may be judged morally bad, because their prudential value outweighs their moral badness. The goal of the following sections is to formulate conditions for such excusable courses of action that allow one to distinguish morally acceptable from morally unacceptable decisions, and for this purpose, thresholds will be introduced. Which threshold is the right one in a given situation, however, is a substantive moral question that I will not attempt to answer in general.

There are some known criticisms that I would like to exclude before going into the details. Decision theory has often been criticized for its purported inability to deal with moral dilemmas. The example of a resistance fighter torn apart between obligations to his country and to his sick mother is often given (Sartre 1946), and another type of such dilemmas may concern long-term choices between different ways of life. If such dilemmas exist, then by definition no account of practical reasoning whose outcome is supposed to be an action, choice of action, or an intention to act can resolve them, and so with respect to such cases practical reasoning theories like those of Broome (2002) and Horty (2012) are in the same boat with decision theory. A genuine moral dilemma has by definition no rational ‘solution’, and since this article is concerned with the morality of decisions in situations in which a rational decision can be made, these types of examples need not be considered further in what follows.

In the remainder of this article, first a brief overview of rational decision making is given in Section 1. Then, in Section 2, conditions are laid out for classifying decisions as moral or not given that morally relevant attributes has been identified. These include unipolar thresholds, then bipolar thresholds are introduced to more adequately represent disvalue, and later the changes needed to deal with uncertainty are discussed. The discussion is rounded up in Section 3 in which the connections of the proposal to existing work in decision making and to utilitarianism are addressed.

1 Rational Choice

The rational decision making tradition and accounts of practical reasoning in moral philosophy have diverged at some point in the history of ideas, and various factors may have contributed to this unfortunate development: What was formerly called Welfare Economics has dropped much of the welfare aspect from its domain of inquiry, the deontic tradition has always been predominant in moral philosophy, and much of the recent work in axiology has focused on special problems of value incommensurability and parity rather than the question of what makes a choice moral.

In order to make this article more or less self-contained, aspects of the decision making perspective on rational choice will be briefly laid in the following paragraphs insofar as they are relevant for the topic of moral decision making, although limitations of space will only allow me to scratch the surface (and there is a dangerous iceberg below). Readers familiar with standard additive decision theory may skip this section. The proposal itself and how it relates to the standard weighted-sum account will be laid out in Section 2.

Among the many topics that cannot be addressed in such a brief survey is the question of the

rationality of decision making principles themselves and Expected Utility Theory in particular. This question has been addressed by a vast variety of authors starting with Ramsey (1931), von Neuman & Morgenstern (1947), Savage (1954), Debreu (1959) and Fishburn (1970). See Eisenführ et. al. (2010) for a modern introduction to applied decision theory, Keeney & Raiffa (1976) for a more technical overview and Bouyssou et al. (2010) for details.

1.1 Decision Making Under Certainty

The key notion of decision theory are preferences. Among several alternatives a decision maker has preferences which reflect his values: One alternative might be better than another or they might be just as good. So to be able to talk about the morality of decisions to act we must assume that moral rules can give rise to corresponding preferences between action alternatives. For example, killing someone for selfish motives is considered murder unless the circumstances are exceptional such as self-defense or wartime. When a decision maker contemplates two hypothetical alternative courses of action and one of them involves murdering someone while the other does not, then avoiding murder should be preferred from a moral point view. In other words, avoiding murder has a higher value than not avoiding it. So does avoiding loss of life in general, but notice that these two alternatives are also comparable; it is commonly presumed that avoiding murder is preferable to avoiding mere loss of life. None of this should be controversial, yet the claim that all kinds of alternatives and aspects thereof are comparable is, of course, not so innocuous and has been attacked from time to time, see for instance the discussion in Chang (1997). But as mentioned above, moral dilemmas that lead to genuine value incommensurability shall not be considered and we focus on the decision-guiding aspects of values. Under this premise, some standard assumptions about the preference relations can be made. First of all, they are complete, i.e., a decision maker either prefers one alternative over the other or considers them equally good.

To get a full-fledged decision theory many more assumptions are needed. Generally, it is presumed that preferences can be described by value functions whose outcome is a numerical value. These numerical values can be compared with each other, reflecting the point of view that the decision maker can decide for any two courses of actions which one has a higher value than the other. In other words, if $a \succsim b$ says 'alternative a is preferred to b or the decision maker is indifferent about the alternatives' (weak preference), then there is a continuous value function $v(\cdot)$ such that $v(a) \geq v(b)$ iff. $a \succsim b$. This fairly strong assumption is also made by subjective expected utility theory of Savage (1954) and must be taken with a grain of salt. It implies that preferences between alternatives are transitive, that there are no principally incommensurable values, and that parity (Chang 2002; Gert 2004; Rabinowicz 2008) can be disregarded as well. Depending on how the values of individual attributes are combined into an overall valuation there are also a number of further, more technical conditions. If they are combined by adding them, the attributes must be mutually preference independent. Other modes of combination such as multilinear models (Keeney & Raiffa 1993), generalized additive independence (Gonzales & Perny 2004), and multiplicative models (Krantz et. al. 1971) impose less strong conditions.

Not all of the standard assumptions about preferences need to be made for choice-guiding actions, for example Fishburn (1991) uses non-transitive preferences and Hansson (2001: Chapter 2)

replaces transitivity by weaker conditions ('top transitivity' and 'weak eligibility'), but for the purpose of this article I will stick to so-called additive models for simplicity. This means that in addition to some more technical conditions, preferential independence and difference independence need to be presumed. These conditions state that a comparison between two attributes does not depend on other attributes. To put it more precisely, preferential independence states that if there are two alternatives a, b that differ only in attribute i and there are two other alternatives a', b' that also differ only from each other in attribute i , and $a'_i = a_i$ and $b'_i = b_i$, then $a \succsim b$ if and only if $a' \succsim b'$.¹ Difference independence is an even stronger condition that says that in the same scenario with two preferences that only differ in one attribute a decision maker will be indifferent between the choice of shifting from a to b and a choice of shifting from a' to b' . If these conditions and a few more technical ones are fulfilled, additive value functions express preferences between alternatives by aggregating the values of their individual attributes. In a model with n attributes, the overall value of an alternative is their weighted sum $v(a) = \sum_{i=1}^n w_i v_i(a_i)$, where w_i represents the relative weight of an attribute (in comparison to the other attributes), v_i is the value function of the i -th attribute and a_i is the value of the attribute. The alternative with the highest value is the one recommended for rational choice; if there are more than one with a highest value, which might of course happen, then the decision maker is indifferent about them and may choose as she likes or attempt to refine the model to decide between these alternatives.

Someone who is less familiar with these kind of models might ask why a value function is used instead of just directly using values for the respective attributes. To motivate this feature, it is instructive to take a glimpse at the related field of Consumer Theory in economics. In this theory each alternative is taken as a bundle of goods, and it is generally assumed that under normal circumstances the more of a good one has in a bundle while the amount of other goods is kept fixed, the higher will be the value of the bundle. Two bananas and an apple are worth more than one banana and an apple. In addition, however, it is often claimed that the Principle of Marginal Utility² holds, which has been confirmed empirically for some domains.³ According to this principle, under normal circumstances the overall value of a good decreases in relation to other goods in a bundle the more units of the good one possesses. Consider the innocuous example of someone who has one banana and is willing to swap it for two apples. If the same person had twenty bananas, she might

¹Other, more technical conditions such as restricted solvability and the Archimedean principle differ slightly between the case with two and three or more attributes but I suppress these details in what follows. Details can be found in Keeney & Raiffa (1993), pp. 104-117, Eisenführ et. al. (2010), pp. 125-155, and Krantz et al. (1971) for the mathematical underpinnings.

2 This principle was first formulated by Gossen (1854) as a characteristic (Merkmal) of all human pleasure (Genuss) in the context of a purely pleasure-utilitarian foundations of economics and later strongly criticized by purporters of the Marxist labor theory of value.

3 See for example Inglehart (1990) for the value of economic growth in societies, Liu (2003) for the quality of marital sex, and Horowitz (2007) for general experimental results.

be willing to swap three bananas for one apple. The value of bananas in comparison to apples has dropped from two times an apple to one third of an apple. To model such cases, *cardinal value functions* are needed instead of purely ordinal ones; in addition to the ordinal information given by qualitative preferences, these represent information about preference intensities and allow for a comparison of preference differences. For simplicity, take the apple value function to be linear, defined by the two points $v_a(1)=0.1$ and $v_a(20)=0.3$. Furthermore, $v_b(1)=0.2$. Following the example, suppose now that $v_b(20) \approx 0.3 \approx \frac{1}{3}v_a(20)$. To ensure that the banana value function does not violate the Principle of Marginal Utility, a third point is needed; suppose $v_b(10) = 0.28$ is this point. Given these premises, the value function may be described by a quadratic function $v_b(x) = 1.2 - x^{-0.0354}$ obtained by curve fitting; it is concave, monotonically increasing, has limit 1.2, and satisfies the Principle of Marginal Utility.

Note that there may be some applications in *moral* decision making with cardinal value functions for which it makes sense to presume this principle and some for which it does not. For example, one might argue that in the evaluation of young delinquent's misdemeanors the difference between the first and the second consecutive petty theft ought to count more than the difference between the twelfth and thirteenth such case. On the other hand, consider the virtue-utilitarian value of polite actions per day such as opening a door for someone else or cheering someone up. Perhaps someone believes that ten such actions a day are optimal but twenty of them are just as good as one. After all, it is possible to be too polite. In that case $v(1)=\frac{1}{4}$, $v(10)=1$, $v(20)=\frac{1}{4}$.⁴ Similar cases also occur in consumer theory, where economists sometimes try to avoid them for technical reasons by only considering the increasing part of the function. I will not presume the principle in general in what follows and the account works for cardinal and ordinal value functions alike.

1.2 Decisions Under Risk

While the focus of this article is on decision making under certainty, some remarks about decisions under risk and uncertainty seem to be appropriate, as decision theory is usually applied in contexts with risk and uncertainty and this might hold in particular for moral decision making. Risk is generally dealt with by presuming the axioms of *Expected Utility Theory* of von Neuman & Morgenstern (1947). Omitting most of the details and the (formal) justification of their approach, an additive decision model under risk can be obtained from the one laid out previously by defining a probability over the alternatives under consideration and then computing the expected utility of the alternatives in a way that is very similar to decision making under certainty. Instead of speaking of a value function it is customary to speak of a utility function that must satisfy the same requirements as a value function. Taking $\vec{a} = \langle a_1, \dots, a_n \rangle$ as a vector of attributes like before, we may define a

4 Again, the function may be approximated, for example by polynomial curve fitting $v(x) = -0.108x^2 + 2.269x - 1.911$ is obtained. It would, of course, also be possible to abstract from the subjective optimum and find a general 'politeness value function' with a parameter for the turning point, but since the example is a bit contrived this matter shall not be pursued further.

consequence as a tuple $\langle p_j, \bar{a}_j \rangle$ ($1 \leq j \leq m$) and take each alternative to have m consequences. The subjective expected utility of an alternative a is then $u(a) = \sum_{j=1}^m p_j \sum_{i=1}^n w_{j,i} u_{j,i}(a_{j,i})$ in such a model.⁵

It is important to bear in mind that under risk a utility function fulfills two roles that are not clearly separable from each other. On one hand, it represents a value function, which might for example exhibit the non-linearity exemplified by the Principle of Marginal Utility. On the other hand, it may also represent the risk attitude of a decision maker. Suppose the decision maker's value function v is linear, i.e. has a straight line as a graph. That means that a change by any positive amount of an attribute has the same value no matter how much of it was already present. For example, getting four bananas is worth the same to the decision maker whether he already has four of them or none of them. If that is the case, a concave utility function represents risk aversion. It is easiest to see this by looking at bets with losses. Consider a 50/50 bet a with a win of 8 apples and a loss of 4. The expected value of this bet is $\frac{1}{2}v(8) + \frac{1}{2}v(-4) = 2$. Take another bet b with the same expected value, say a 50/50 bet with win of 20 apples and loss of 16: $\frac{1}{2}v(20) + \frac{1}{2}v(-16) = 2$. Concavity of the utility function implies that $u(a) > u(b)$. Hence, the decision maker is more willing to take bet a with a possible loss of 4 apples than the more risky bet b with a possible loss of 16 apples. A similar argument shows that a convex utility function implies that the decision maker is risk prone *unless* he already he already exemplifies the Principle of Diminishing Marginal Utility to a very high extent.⁶ As laid out by Schoemaker (1982), the fact that the two functions of expected utility, subjective valuation versus risk attitude, are often not clearly kept apart in the literature has led to confusions. The following sections will focus on decision making under certainty and therefore avoid these problems.

2 Moral Decision Making

Let us return to the initial question of how to classify decisions as moral in a way that does not simply give moral absolute priority over other considerations. The goal is to find conditions for what makes a decision moral without taking *only* moral aspects of a decision situation into account. It seems obvious that such an approach must involve a threshold at one place or another; the decision maker is allowed to deviate from moral rules to some extent, but not when it matters and not too much.

Before going on, a flawed account needs to be addressed, because it might seem intuitively appealing at first glance. What about the claim that a decision maker ought not deviate from the prescriptions of underlying moral rules 'too often'? This is not a good option. A highly immoral decision maker may make moral decisions most of the time. What matters in the end is what is at

5 The notation and terminology differs slightly from that of Savage (1954), whose setting is more general than the one presented above. For what follows, the differences do not matter.

6 Arguably, it is harder to find examples of this attitude than in the converse case of diminishing marginal utility.

stake with each decision and not how often similar decisions over the same types of action are made. In an overall consequentialist approach what is at stake is determined by potential wins and losses, and a theory that discriminates between different stakes will at some point involve either thresholds or special functions over the outcomes. The threshold view is simpler and, as I believe, the adequate way of making the moral decision process permeable and ‘soft’.

2.1 The Unipolar Threshold View

If a moral component can be distinguished from a non-moral component in decision making in a given situation, then certain attributes of alternatives must be morally relevant while others are not. What counts as morally relevant and not is a matter of the underlying moral theory. In a broadly-conceived deontic setting the morally relevant attributes are those governed by a moral rule or norm. For example, the attribute ‘number of lives lost’ of a consequence of some alternative course of action is morally relevant, because notwithstanding certain exceptional circumstances minimizing loss of life might be considered a moral obligation or one’s duty. In contrast to this, the attribute ‘pleasure of taste’ concerns what von Wright (1963) calls a hedonic good and from a deontic perspective does likely not have moral significance.⁷ It falls into the category of prudential value. On the other hand, from the perspective of a classical utilitarian, hedonic goodness might very well be morally relevant for its contribution to the overall welfare of a group. This difference illustrates the dependence of moral relevance on the underlying moral theory, and the precise nature of this connection is an open problem. Moreover, it seems often reasonable to assume that one and the same attribute can be morally relevant in one instance and irrelevant in another, and this issue is closely related to the previous one. Since it would go far beyond the scope of this article to address these problems here, I will assume in what follows that certain persons with moral expertise, such as perhaps ethics committees, can decide between morally relevant and irrelevant attributes in a given choice situation.⁸

Under this assumption, let us write M to denote the set of morally relevant attributes and N for the set of other attributes. A table that depicts the relevant parts of a value function over alternatives and attributes will from now on be called a *decision table*. Such a table may be said to *rationaly ground* a decision if the decision maker acts according to it. Only rationally grounded decisions are considered from now on.

To tackle the problem we are dealing with, let me first introduce the concept of *dominance* and then lay out a similar concept for moral attributes. An alternative a dominates an alternative b if and only

7 There might be a rule, however, stating that no one is permitted to needlessly deprive someone else of personal pleasure. However, it seems striking that this must be based on some threshold as well, since there are many situations in which it is customary and legitimate to deprive people of pleasure—think for instance about work, which is rarely *always* fun. Notice further that methods for finding a balance of power and distinctions like that between positive and negative rights might be needed.

8 Which account of morality is the right one is a decidedly moral question for anyone but an extreme moral relativist.

if $w_i v_i(a_i) > w_i v_i(b_i)$ for all i and for at least one j , $w_j v_j(a_j) > w_j v_j(b_j)$. Dominated alternatives can be discarded, since at least one alternative is always preferred to them. The idea is now to introduce a similar concept for moral versus other attributes. A decision matrix is *fully moral* if and only if for all alternatives a, b the following condition holds:

$$\text{If } \sum_{a_i \in M} w_i v_i(a_i) > \sum_{b_i \in M} w_i v_i(b_i), \text{ then } v(a) > v(b). \quad (1)$$

A decision maker whose decisions always satisfy (1) never makes any moral mistakes, is a perfect moral decision maker and perhaps also a moral tyrant in the sense laid out above. There is something eerily wrong about such a person. Since the antecedent holds for arbitrary alternatives, non-moral values only play a role in her decision if the weighted sums of the moral attributes in the antecedent are exactly equal, i.e., if the alternatives have exactly the same moral consequences. Otherwise they can be discarded. The moral attributes absolutely dominate the non-moral attributes.

It might be tempting to reply that a decision maker acts morally as long as *most* decisions satisfy (1), but as mentioned in the beginning of this section, this kind of reply is not acceptable without further elaborating the role of the stakes. A murder cannot be excused by the fact that the murderer acts morally most of the time, even though this fact might count in favor of him in court. So it seems that the condition must be relaxed in another way. One solution is to stipulate, as an additional constraint, that the sum of the moral attributes of the first alternative in the antecedent must be higher than a certain threshold. In other words, we are looking for conditions to further narrow the set of attributes that is morally relevant in general down to a smaller set of attributes that are morally relevant in a particular decision situation based on the values of the attributes in question. A decision matrix is *moral relative to global threshold α* if and only if the following condition holds for all alternatives a, b :

$$\text{If } \sum_{a_i \in M} w_i v_i(a_i) > \alpha \text{ and } \sum_{a_i \in M} w_i v_i(a_i) > \sum_{b_i \in M} w_i v_i(b_i), \text{ then } v(a) > v(b). \quad (2)$$

However, this condition will only work as desired, as long as it can be ensured that any value of an attribute or combination of attributes that is considered morally relevant exceeds the threshold. Not only is it hard to conceive how a reasonable moral theory could provide such a threshold, it would also be problematic to elicit such a joint feature of attributes from a person or expert panel. It seems more reasonable to stipulate thresholds for individual attributes instead. A decision matrix is *moral relative to thresholds α_i* if and only if the following condition holds for all alternatives a, b :

$$\begin{aligned} \text{If there is an } \alpha_i \text{ s.t. } w_i v_i(a_i) > \alpha_i \text{ and } \sum_{a_i \in M} w_i v_i(a_i) > \sum_{b_i \in M} w_i v_i(b_i), \\ \text{then } v(a) > v(b), \end{aligned} \quad (3)$$

where α_i is the threshold of the i -th attribute a_i . If just one moral attribute exceeds the threshold, then the whole set of moral attributes becomes relevant.

This principle seems to reflect the way in which a person's actions are often judged retrospectively, for example in court or public opinion, and incorporates a certain virtue-ethical stance. For example, when someone is accused of a crime, judges and plaintiffs sometimes take into account moral aspects of additional motives, whether for instance the crime has been committed out of need or sheer selfishness. The goal is hereby to determine the character of the accused and the final verdict hinges to some extent on this assessment. Despite being common practice, this method seems questionable for the present, more general purpose. If a threshold plays the role of determining moral relevance, it seems that an attribute below the threshold ought not enter moral considerations, or otherwise it is no longer clear what the thresholds actually does.

By varying the condition slightly, a more appropriate evaluation method can be obtained. Let T be the set of thresholds in the model indexed in the same way as the set of moral attributes, and let M_a^T denote the set of indices i of attributes of alternative a in M such that $w_i v_i(a_i) > \alpha_i$. The antecedent condition is then relativized to this set. A decision matrix is *selectively moral relative to a set of thresholds T* if and only if the following condition holds for all alternatives a, b :

$$\text{If } M_a^T \neq \emptyset \text{ and } \sum_{i \in M_a^T} w_i v_i(a_i) > \sum_{i \in M_b^T} w_i v_i(b_i), \text{ then } v(a) > v(b). \quad (4)$$

In this condition, only moral attributes whose values exceed their respective threshold are considered relevant in a specific decision situation. Once these attributes have been identified, they are weighed against other moral attributes in the set.

To get an idea how this condition works consider the following trivial example. Suppose Bob is contemplating whether he should steal his cousin's chocolate bar and suppose furthermore that he must not fear reprimand, because his cousin does not keep track of his huge inventory of chocolate bars. Let feature 1 be an artificial measure that is higher when no chocolate is stolen.⁹ On the other side of the equation is Bob's personal pleasure, represented by feature 2. For simplicity all weights are taken to be 1 in this and the following examples, and the threshold of the moral attribute 1 in this example is $\alpha_1=0.3$. Suppose the following matrix represents his decision:

	a	b
1 – not steal	0.4	0.2
2 – personal pleasure	0.2	0.6

⁹ A better way to model this situation will be laid out in the next section. For the time being, you may for example, assume that any value above 0.38 indicates that no good has been stolen.

Moral sum:	<u>0.4</u>	0.2
Total sum:	0.6	<u>0.8</u>

As a rational decision maker, Bob decides to steal the chocolate bar. Determining whether his decision is morally permitted is straightforward. First, mark all rows with moral attributes. This is the first row in this example. Compute their ‘moral sum’ in each column and underline the highest values in the moral sum row. In this case the winner is *a* with value 0.4. Second, compute the total sum and underline the highest values in that row, too. In this case, the winner is *b* with value 0.8. Evaluation: If a value is underlined in a total sum column and not in the corresponding moral sum column, then the decision is non-moral; otherwise it is moral. This is just in ordinary terms what Condition (4) says. Using this procedure, it becomes apparent that Bob’s decision in the above example is non-moral. In virtue of being non-moral, the decision is also immoral in this case, because a morally preferable course of action was available and could have been taken.

One might think that there is an easier way to evaluate such an example. Instead of summing up the values of moral attributes, as condition (4) prescribes, one might replace the antecedent by individual comparisons, i.e. all moral attributes must satisfy $w_i v_i(a_i) > w_i v_i(b_i)$ in the antecedent. This modified principle would, however, predict that *any* action based on the following matrix is immoral:

	a	b
<i>1</i>	0.7	0.2
<i>2</i>	0.3	0.8
Moral sum	<u>1.0</u>	<u>1.0</u>
Total sum	<u>1.0</u>	<u>1.0</u>

In this example two different moral attributes outweigh each other. Consequently, any of *a* or *b* ought to be just as good from a moral point of view, which is correctly predicted by Condition (4) and the corresponding informal evaluation procedure outlined above. The shortcut version does not make the correct prediction here, as it only takes into account the relations between individual moral attribute comparisons and the total sum.

A final abstract example illustrates a mixed case with three alternatives and motivates our talk about decision tables as opposed to actual decisions. These tables encode additional information that may sometimes turn out to be useful for an assessment.

	a	b	c
<i>1</i>	0.7	0.2	0.4

2	0.3	0.7	0.3
3	0.2	0.5	0.7
Moral sum	0.9	0.7	<u>1.1</u>
Total sum	1.2	<u>1.4</u>	<u>1.4</u>

The rule predicts that a decision based on this table is immoral, because there are two highest values in the total sum row, but only row *c* has a corresponding highest moral sum. However, if the decision maker were to *actually choose* alternative *c* his choice would be morally impeccable on purely consequentialist grounds. Still, the underlying table is at least problematic from a moral point of view, as he could just as well have chosen alternative *b*. Even if the decision maker actually chooses *c*, he might not have done so for the proper motive. For example, he could have made the choice by throwing a coin. This example shows that moral applications of decision theory need not be solely consequentialist in nature even when they involve the weighing of alternatives that represent different possible courses of action.

2.2 The Bipolar Threshold View

The above way of laying out decision problems is clumsy to say the least. In the first example, an artificial attribute ‘not steal’ is used to express the moral value of not stealing instead of expressing the *disvalue* of stealing something. The reason for this was that the thresholds were formulated for positive values. If a moral attribute’s value exceeded a threshold, the value was considered morally relevant. Simple additive models may also include negative values, but then the threshold view must be adjusted. The resulting model is *bipolar*, as it distinguishes between negative and positive values. Let the set M_a^T contain index i if and only if (Case 1) α_i is a positive threshold and $w_i v_i(a_i) > \alpha_i$, or (Case 2) α_i is a negative threshold and $w_i v_i(a_i) < \alpha_i$. Apart from that no further changes are needed and Condition (4) remains intact.

With this adjustment, disvalues with a corresponding negative moral thresholds can be used. Although the change is minimal, its conceptual relevance is huge, as now the theory may be used to express positions like *Negative Utilitarianism*. For example, one might believe that it ought to be allowed to cause small amounts of harm to persons without the harm becoming morally relevant. Everyone does this almost every day, for example when arguing or having a bad day, and people are not always harmed for the sake of a greater good. It would amount to moral tyranny to consider any kind of harm done to a person morally reprehensible. After all, not many executives of a company can do their job properly without occasionally causing displeasure among their employees. What matters is whether the amount of pain exceeds a certain threshold, which is described by the bipolar threshold view. Negative Utilitarianism has been attacked by Smart (1958) and recently by Ord (2013), who gives an excellent overview of the main arguments against it. While the details of this debate are beyond the scope of this article, it is noteworthy that this particular version of what Ord calls ‘Lexical Threshold View’ fares better than most other variants of negative utilitarianism. Ord considers the sudden change in evaluation once a threshold is reached implausible for small,

perhaps even arbitrary increments of disvalue and constructs a form of Sorites paradox against it (his ‘Continuity Argument’). I believe Ord’s argument does not speak conclusively against the above version of the threshold view but would like to leave this matter for another occasion.¹⁰

2.3 Dealing with Stakes

Let me finally address the role of stakes in decision making under risk and uncertainty very briefly. The bipolar threshold view does not require many modifications in this setting. Basically, $w_i v_i(a_i)$ must be replaced by $p_j w_{i,j} u_{i,j}(a_{i,j})$ in the above conditions. The threshold is applied to the weighted outcome times the probability of the occurrence of a consequence. Although Savage (1954) has shown, within a different yet sufficiently similar setting, that some intuitively plausible principles governing preferences and subjective plausibility imply that the factors $\sum_{j=1}^m p_j$ indeed constitute , and so subjective expected utility theory as a whole is derivable from few, intuitively compelling principles, there are legitimate concerns about moral implications of always using expected utility. What is sometimes treated as a risk might in reality be based on some form of epistemic uncertainty. In that case, the Precautionary Principle might dictate that maximal losses ought to be minimized. According to this so-called *Maximin* method, instead of multiplying a weighted utility with a probability, the alternative with the smallest loss is the most preferable one. It seems that this method could be justifiable for moral attributes on moral grounds alone, as long as the risk in question really is uncertainty in disguise.¹¹ There are other decision principles such as *Minimax with Regret* to consider if the stakes are high and some form of uncertainty is at play. It would go beyond the scope of this article to enter this debate, but I cannot see any principal obstacles that would keep us from adjusting the bipolar threshold view to such alternative decision principles, and conditions like (4) seem to be applicable to these as well.

3 Concluding Remarks

Several conditions for evaluating the morality of a decision on the basis of given sets of morally relevant and non-relevant attributes and corresponding thresholds have been investigated in a (simplified) decision-theoretic setting. Among these the bipolar threshold view of Condition (4) turned out to be the most general and adequate. Once moral attributes and their thresholds have

¹⁰There is a strong argument against any view that rejects lexical threshold utilitarianism while endorsing classical utilitarianism. From a purely formal point of view it seems possible to translate a threshold model into the classical approach by choosing suitable utility functions of potentially unusual shape. If that line of thinking is correct, as I believe it is, then thresholds are merely a convenient way of making decision boundaries explicit, which is what makes them useful for the purpose of moral decision making.

¹¹One might argue that this problem just concerns the possibility of incorrect modeling, yet prescriptive theory ought to be based on the assumption that the modeling is correct. However, if the stakes are high enough, a ‘Meta-Precautionary Principle’ might no longer be clearly discernible from the normal one.

been identified, this condition could be used to give moral advice. For example, in a group decision making process a team of ‘moral experts’ could provide the value functions and thresholds for moral attributes, which may then be combined during an argumentative decision-making phase with the outcome of a preference elicitation process of another group consisting of experts about the specific application domain such as public health professionals.

Although this suggestion must be taken with a grain of salt in light of the simplifying assumptions mentioned in Section 2, it is worth noting that most of the known criticism of these assumptions apply to decision theory in general and have been addressed in great detail in the seminal literature already. For instance, worries about preferential independence, completeness and transitivity of the underlying preferences are addressed by Fishburn (1991), Hansson (2001) and in contributions to Bouyssou et al. (2010), and a number of alternatives to the additive decomposition of value functions such as multilinear models, multiplicative models, and generalized additive decomposition have been on the table for a long time. Within this spectrum, the present account is a lexicographic outranking method, mixed with additive models for simplicity.

Another kind of criticisms based on the fact that we do not make decisions in the way prescriptive decision theory mandates are also well known, see for instance Kahneman & Tversky (1979; 2000), arguments by Broome (1999: Ch. 6) for Bolker-Jeffrey utility theory, and the Maxmin and Minmax with Regret approaches mentioned above. As suggested in the last section, it seems possible to adjust the bipolar threshold view to such alternative frameworks, but perhaps this is not needed. Some of the approaches mentioned above, such as work by Kahneman and Tversky and work in the evolving field of ‘ecological rationality’, draw their motivation from empirical aspects of decision making, and there are doubts whether these successfully undermine the normativity of expected utility theory that is established by intuitively plausible rationality postulates. I am personally wary of any account of rationality whose justification is mainly derived from empirical success criteria. Be that as it may, the matter seems to be undecided even among scholars of decision theory.

There is another worry about the threshold view that has to be mentioned. Thresholds constitute sharp boundaries that in practice might make the approach unfruitful. If thresholds for moral attributes are always so low that the cases when they are not relevant are utterly trivial, then the differences between (1) and (4) might become uninteresting. Perhaps the stakes are always high in situations worthy of thorough decision analysis. Whether this is a problem or not can only be decided in practice.

Finally, one might wonder whether the theory laid out so far could serve as a basis for a utilitarian calculus. The answer is clearly no. First, there are elements of virtue ethics in the account; not only the actual decision counts for an assessment, but the whole decision table on the basis of which it has been made. Second, a full-fledged utilitarian calculus requires much more than what multi-criteria decision theory can offer. Time has to be taken into account, for there is no doubt that many people prefer short amounts of intense displeasure (pain, inconvenience, dissatisfaction) over a long period of lesser suffering, and, vice versa, are sometimes willing to tolerate extended amount of displeasure—morally acceptable variants of the so-called ‘necessary evil’— to later obtain a greater good. If at all, the bipolar threshold view can only become a small part of such an approach,

allowing one to consider single decision situations from a moral aspect provided that the relevant alternatives including their consequences over time can be identified and an appropriate connection between attribute values, their thresholds, and moral rules can be drawn. How such a connection would look like and how it fares with known methods of preference elicitation is a surprisingly open question of moral philosophy, though.

Third, while decision making can account for the influence of possible courses of actions on other people, the simple version laid out above cannot represent equilibria between preferred choices of several decision makers. Such equilibria between morally relevant values would have to be the cornerstone of a full-fledged utilitarian calculus and have not been addressed at all. As is well-known, there are a number of obstacles to such an ambitious project, ranging from the question of how to ‘tame’ respective variants of Arrow’s theorem to more philosophical worries about value aggregation and the interpersonal comparability of utilities. It is also well-known that Pareto optimality used in economics allows for social states that are inherently unjust, and so a utilitarian calculus worth to be taken seriously would have to take into account additional justice criteria. Variants of decision theory like the one laid out above may help in making morally acceptable decisions, but in a social context only on the basis of an existing theory of justice. For itself, decision theory does not contribute to such a theory and also cannot provide the criteria for deciding at what level attributes become morally relevant.

References

Broome, John (1999): *Ethics out of Economics*. Cambridge University Press.

Broome, J. (2002), ‘Practical Reasoning’. In Broome, J.: *Reason and Nature: Essays in the Theory of Rationality*, Oxford University Press, .

Chang, R. (1997), ed.: *Incommensurability, Incomparability, and Practical Reason*, Harvard University Press, .

Chang, R. (2002): ‘The Possibility of Parity.’ *Ethics*, Vol. 112, p. 669-688.

Debreu, G. (1959), *Theory of Value: An Axiomatic Analysis of Economic Equilibrium*, Yale University Press, New Haven, London.

Eisenführ, F., Weber, M. & Langer, T. (2010): *Rational Decision Making*. Springer.

Fishburn, P. C. (1970): *Utility Theory for Decision Making*. John Wiley & Sons.

– (1991): ‘Nontransitive additive conjoint measurement.’ *Journal of Mathematical Psychology*, Vol. 35, p. 1-40.

Gert, J. (2004): ‘Value and Parity.’ *Ethics*, 114, p. 492-510.

Gossen, H. H. (1854): *Entwicklung der Gesetze des menschlichen Verkehrs und der daraus fließenden Regeln für menschliches Handeln*. Vieweg.

Hansson, S. O. (2001): *The Structure of Values and Norms*. Cambridge University Press.

Hilpinen, R. & McNamara, P. (2013). ‘Deontic Logic: A Historical Survey and Introduction.’ In

- Gabbay, D. ; Horty, J. ; Parent, X. ; van der Meyden, R. & van der Torre, Leendert: *Handbook of Deontic Logic and Normative Systems*. College Publications.
- Horowitz, J., List, J. & McConnell, K. E. (2007): 'A Test of Diminishing Marginal Value.' *Economica*, Vol. 74, p. 650-663.
- Horty, J. F. (2012), *Reasons as Defaults*, Cambridge University Press, Oxford / New York.
- Inglehart, R. (1990): *Culture Shift in Advanced Industrial Society*. Princeton University Press.
- Kahneman, D. & Tversky, A. (1979): 'Prospect Theory: An Analysis of Decision Under Risk', *Econometrica*, Vol. 47, 263-291.
- Kahneman, D. & Tversky, A. (eds.) (2000): *Choices, Values and Frames*. Cambridge University Press.
- Keeney, R. L. & Raiffa, H. (1976, 1993), *Decisions with Multiple Objectives: Preferences and Value Tradeoffs*, John Wiley & Sons, New York. Cit. in 1993 edition published by Cambridge University Press.
- Krantz, D. H.; Luce, R. D.; Suppes, P. & Tversky (1971), *Foundations of Measurement*, Vol. I, Academic Press, New York.
- Liu, C. (2003): 'Does Quality of Marital Sex Decline With Duration?' *Archives of Sexual Behavior*, Vol. 32, No. 1, p. 55-60.
- Meinong, A. (1894). *Psychologisch-ethische Untersuchungen zur Werttheorie*. Leuschner & Lubensky.
- von Neuman, J. & Morgenstern, O. (1947): *Theory of Games and Economic Behavior*. Princeton University Press.
- Von Wright, G. H. (1963): *The Varieties of Goodness*. Routledge & K. Paul.
- Ord, T. (2013), 'Why I'm Not a Negative Utilitarian', University of Oxford, published online at URL <http://www.amirrorclear.net/academic/ideas/negative-utilitarianism/index.html>. Date of retrieval: 20. 2. 2014.
- Rabinowicz, W. (2010): Value Relations – Old Wine in New Barrels, In: Reboul, A. (ed.), *Philosophical Papers Dedicated to Kevin Mulligan*. Université de Genève. Publ. Online. URL <http://www.philosophie.ch/kevin/festschrift/>.
- Ramsey, F. P. (1931), 'Truth and Probability'. In Ramsey, F.P.: *The Foundations of Mathematics and other Logical Essays*, Kegan, Paul, Trench, Trubner & Co., London, pp. 156-198.
- Sartre, J.-P. (1946), *L'existentialisme est un humanisme*, Éditions Nagel, Paris.
- Savage (1954): *The Foundations of Statistics*. Wiley.
- Schoemaker, P. J. H. (1982): 'The Expected Utility Model: Its Variants, Purposes, Evidence and Limitations.' *Journal of Economic Literature*, Vol. 20 (June 1982), p. 529-563.

Smart, R. N. (1958), 'Negative Utilitarianism', *Mind* **67**, 542-3.