

Dealing with Moral Uncertainty: Do Logical Properties Help?

Wulf Gaertner^{1,2}

¹Department of Economics, University of Osnabrück, Osnabrueck, Germany

²CPNSS, London School of Economics, London, UK

Email: wulf.gaertner@uni-osnabrueck.de

How to cite this paper: Gaertner, W. (2021). Dealing with Moral Uncertainty: Do Logical Properties Help? *Open Journal of Philosophy*, 11, 1-15.
<https://doi.org/10.4236/ojpp.2021.111001>

Received: October 24, 2020

Accepted: December 22, 2020

Published: December 25, 2020

Copyright © 2021 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).
<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

If an agent is unsure about which moral theory or principle should guide her action in a decision situation, she faces moral uncertainty. In recent years, various strategies have been explored to deal with this type of uncertainty. In this paper, we briefly mention two strategies from the literature that make use of credence distributions over moral theories, namely “my favourite theory” and “maximizing expected choice-worthiness”. As an alternative, we propose a two-step procedure which uses the concept of aggregation over structural properties. It is standard in the theory of collective choice and has recently been applied to the Kuhnian problem of theory choice as well. The idea is to explore how a morally-motivated rational agent may assess different moral theories on the basis of fundamental properties. These properties are ranked on a common scale of qualitative verdicts. This method enables an outside observer to make comparisons across competing moral theories and then conclude to what degree these theories fulfil a set of postulated properties. By doing so, we try to render the reasons more transparent that lie behind different types of credence ascriptions.

Keywords

Moral Uncertainty, “My Favourite Theory”, “Expected Choice-Worthiness”, Scoring Functions, Structural Properties, Common Scale of Qualitative Verdicts

1. Introduction

Moral uncertainty describes a situation in which a decision-making agent is unsure about what the correct guiding moral principle or theory is. One might find Rule Utilitarianism, Virtue Ethics, and Kantianism plausible but be certain of none of them. One might even go so far as to argue that moral principles are to a

certain degree context-dependent, dependent, for example, on circumstances such as extreme needs or cultural constraints. Note that moral uncertainty is not reducible to uncertainty about features of actions. One can face moral uncertainty even if one knows all the details of the features of one's actions, including its consequences (see [Sepielli, 2013: p. 580](#)).

A variety of strategies for dealing with moral uncertainty have been explored in the literature ([Gustafsson & Torpman \(2014\)](#) provide a good overview). Two approaches stand out, namely “my favourite theory” (MFT) and “maximizing expected choice-worthiness” (MEC). Under “my favourite theory”, an agent is supposed to choose the action which is permitted by the theory which she favours most ([Gustafsson & Torpman, 2014: p. 159](#)) or, as [Gracely \(1996: p. 328\)](#) stated: “The proper approach to uncertainty about the rightness of ethical theories is to determine the one most likely to be right, and to act in accord with its dictates”. According to “maximizing expected choice-worthiness”, given that one has a credence distribution over the considered moral theories and that the choice-worthiness of an action is cardinally measurable within a theory and comparable across theories, the action which has the highest expected choice-worthiness should be chosen ([MacAskill, 2014: p. 20](#)), for a more elaborate discussion of these two approaches, see [Gaertner & Wüthrich \(2020\)](#).

In this paper, we replace the idea of assuming or postulating credence distributions over moral theories by a proposal that formulates a two-step procedure. This process makes use of an aggregation mechanism which has been widely discussed in the theory of social choice and has recently been applied to the problem of theory choice as well ([Gaertner & Wüthrich, 2016](#)). This approach tries to make the reasons that motivate the ascription of credences more transparent and has a number of desirable properties that, in our view, credence-based theories do not offer.

Our proposal fits into a broader development in decision analysis which has been recently described as an “argumentative turn” ([Hansson & Hirsch Hadorn, 2016](#)). The turn highlights and refines existing tools for dealing with various types of uncertainty other than expected utility maximisation.

The paper is structured as follows. In *section 2*, we introduce the basic features of our alternative approach to handle moral uncertainty. In *section 3*, we lay out in detail our two-step procedure. The concluding *section 4* highlights desirable features of our alternative proposal.

Let us be clear from the outset about the nature of our approach to moral uncertainty. We are not aiming at providing a descriptively accurate picture of the psychology behind decision-makers who face moral uncertainty. Undoubtedly, beliefs in the strengths of respective moral theories play a role in such descriptive accounts. What we aim at is to propose an alternative approach of grappling with situations of moral uncertainty. Credences in our proposal are not based on probability distributions over moral theories but are grounded on structural properties that alternative moral theories possess or should possess if a normative position is taken.

2. A Novel Approach to Decision-Making under Moral Uncertainty

Let us clarify the status of this section. *In principle*, our non-credence based set-up can be viewed as complementary to MFT and MEC, namely as an algorithmic procedure to resolve the problem of moral uncertainty. Admittedly, our scheme may not hold in all cases since the requirements in relation to (individual) rationality may be too demanding in some situations. We start by laying out the basic idea and its structure. Our concrete proposal is then presented in section 3.

Our proposal should primarily be viewed as a normative approach and empirical research would have to show whether in the process of finding the right answer (i.e. action) in a given situation, our suggestion would have constructive virtues. The reader will see, however, that in our argumentation we shall not totally ignore rather concrete situations since we believe that our normatively oriented approach can be of at least some guidance for conflicting situations in real life. One should not forget that MEC also makes high demands of rationality. Deriving a probability distribution over alternative moral theories presupposes that the individual concerned disposes of a broad spectrum of experience, namely alternative (controversial) cases without which the formation of a credence distribution is hardly possible. Much less demanding in terms of rationality requirements were situations in which individuals follow an ingrained moral principle which becomes their favourite theory. However, what would happen if these individuals were at one point confronted with a totally different environment in which their favourite moral would hardly be applicable anymore?

Instead of merely considering the plausibility that a decision-maker assigns to alternative moral theories and to the degree of choice-worthiness that these theories assign to different options, we shall look at fundamental properties that constitute those alternative moral theories over which we wish to gain more insight and clarity. In this way, we hope to render the search for a satisfactory moral theory more objective, at least more intelligible. We think that a decomposition into basic properties that underlie alternative moral theories will make similarities and dissimilarities among these theories more transparent and, hopefully, more comparable. Given this, our approach defends a non-instrumental perspective of the content of moral theories. According to this view, “an accurate representation of a moral theory should capture not only the theory’s deontic content, but also the underlying reasons or principles” (Dietrich & List, 2017: p. 423).

To clarify our own position, we offer an analogue that refers to social choice theory. The majority rule can be viewed in terms of its mechanical scheme where pros and cons in relation to a set of alternative objects are registered and added. This mechanism can further be related to the fact that a single person (a pivotal voter in the jargon) can tip the balance in favour of either of the given alternatives. However, one can also study and evaluate this rule with a focus on the un-

derlying properties which are an equal treatment of voters and alternatives to be decided upon and the property of (a high degree of) responsiveness which reflects the fact that viewed from a status quo, let's say, smallest changes in the underlying preference profile of the voters may result in totally different outcomes, in contrast to an absolute majority rule, for example. Our position is not to consider how the majority rule "performs" in relation to certain issues in a given parliament or in relation to national referenda (as in the UK referendum in 2016) but to consider whether and under which conditions the majority rule is the appropriate scheme or measure in relation to alternative issues, given the underlying properties as described above.

Assume that a single agent is facing a particular decision problem under moral uncertainty where she has to choose among n (with $n > 1$) options for acting. Assume further that she considers m (with $m > 1$) moral theories to be prima facie relevant for the problem at hand. In a first step, the agent evaluates the moral theories as such. This is done via a number of theory evaluation criteria that we consider as basic. These criteria will be subjected to some grading procedure. The output of this process is a set of relevant moral theories. We cannot expect that this set will necessarily be a singleton. In a second step, the agent evaluates the relevant moral theories according to a set of substantive content criteria to identify, if possible, a *single best theory*. The agent is supposed to act according to the advice of this single best theory. We argue that the concept of a scoring function that weighs and compares alternative options is a suitable tool to perform the evaluations in the first and second step.

Our approach may remind the reader of [MacAskill \(2014, 52ff\)](#) who suggests to apply the Borda rule in cases in which the moral theories under consideration do not give a sense of the magnitude of choice-worthiness or in cases in which the magnitudes of choice-worthiness cannot be compared across theories. The main difference between [MacAskill's \(2014\)](#) approach and ours is that [MacAskill \(2014\)](#) has alternative moral theories assign scores to actions while our approach has evaluation criteria or fundamental properties assign scores to candidate moral theories.

From our point of view, [MacAskill's \(2014\)](#) proposal has two weaknesses. First, the Borda rule itself is rather rigid in the sense that it orders options in a strictly linear and an equidistant way so that the articulation of preference intensities is rather limited. Accordingly, it rules out the possibility that scores or ranks which are attached to alternative options manifest smaller or larger gaps amongst each other. As we argue in more detail below, we take this to be too restrictive for the case of handling moral uncertainty. Second, the approach is not fine-grained enough in a further sense. [MacAskill's \(2014\)](#) credence-weighted Borda rank method is directly applied to assess alternatives. Hence, it does not treat the question of how to evaluate theories as a separate task. In fact, he considers this problem to be accurately captured by the credence distribution over moral theories.

3. A Two-Step Procedure

3.1. Step 1: Identifying Relevant Moral Theories

We take it to be the case that credences in moral theories should track plausibility judgments about these theories. The first step of our procedure suggests a way to make the reasons behind these plausibility judgments explicit. Accordingly, this step aims at expressing the conceptual basis behind credence distributions over moral theories in a more lucid way. The idea is to confront alternative moral theories with categories of criteria that may be arranged hierarchically, but not necessarily so.

Consider the following example well known in philosophical discussions. Anna has to make a choice between 1) pushing the fat man on the track, thereby derailing a trolley, and saving five people, and 2) doing nothing. She considers the following moral theories to be *prima facie* relevant: Utilitarianism, Kantianism, and Virtue Ethics. Importantly, this judgement of *prima facie* relevance does not involve a credence distribution over these three theories. Anna is now supposed to evaluate these three theories in light of two categories of criteria: a) general theory evaluation criteria, and b) content criteria. Here are some examples of salient criteria¹:

a) *General theory evaluation criteria*: internal consistency, fruitfulness, scope, simplicity, action-guidance;

b) *Content criteria*: an equal treatment of all humans; integrity and inviolability of the human body.²

We are not claiming that the lists for both criteria are exhaustive. In fact, we do not want to and probably cannot specify a fixed list of criteria across all potential decision problems. The choice of criteria can be context-sensitive in this sense. This is particularly the case for the content criteria which might vary considerably given different decision problems at hand (e.g. a bioethical decision versus a decision in development economics). We will say more about the context-dependency of the content criteria below. In this first evaluation step, Anna is supposed to assess the competing moral theories solely according to the general theory evaluation criteria.

Let us say more about the proposed general theory evaluation criteria: A moral theory is *internally consistent* if it is free from logical contradictions. An ethical theory is *fruitful* if it can be applied to new problems (e.g. moral challenges that arise in the context of a technological innovation or fairness consid-

¹The general theory evaluation criteria are borrowed to a large extent from Kuhn's classical discussion of theory choice (see Kuhn, 1977).

²Cuneo and Shafer-Landau (2014: p. 405) discuss a similar list under the name of "moral fixed points". These include for example the propositions "it is *pro tanto* wrong to engage in the recreational slaughter of a fellow person" and "it is *pro tanto* wrong to impose severe burdens on others simply because of their physical appearance". The appropriateness to use content criteria (as opposed to solely rely on formal criteria) to evaluate moral theories can also be substantiated by a recent insight put forward by Dietrich and List (2017). They show that moral theories can be characterised by two parameters (i.e. which properties or objects of moral choice matter and how a specification of these properties matter). Both of these parameters are content-related parameters.

erations in socio-political contexts). The *scope* of a theory indicates the range of application of the theory. *Simplicity* can be spelled out in terms of amount of guiding principles, interactions between guiding principles, and cognitive accessibility of a moral theory. *Action-guidance* indicates whether a moral theory has an internal method which allows the ranking and comparison of options (Gracely, 1996: p. 329).

Assume further that each of the general theory evaluation criteria provides an ordinal ranking of the moral theories under consideration. The situation in our example is in **Table 1** at the bottom of this page.

How should we go about aggregating the information provided by these rankings?³ We suggest to introduce a scoring rule that was proposed by Gaertner and Xu (2012). This scoring rule has some similarity to the Borda rule but is embedded in a cardinal structure of grades or ranks. In a nutshell, the scoring rule works as follows. Similar to the Borda rule, each criterion ascribes a score to each of the alternatives, and the overall ranking is determined by the overall score of an alternative. However, in contrast to the Borda rule, each criterion does not necessarily generate a linear ordering over the objects or actions to be evaluated but allows for instances of indifference or equivalence as well. The latter implies that if there are, for example, n grades and an equal number of objects to be evaluated, some of the available grades or scores may not be assigned in certain cases so that there are “empty cells”, so to speak. This marks a major difference to the Borda rule. More precisely⁴, let X be the universal set of moral theories containing a finite number of elements. Let N be the set of criteria deemed relevant with $n > 1$. Let $E = \{1, \dots, E\}$, with the cardinality of this set being larger than one, be a set of given positive integers from 1 to E . These integers will in most cases be assumed to be equally distanced.

A scoring function $s_i : X \rightarrow E$ is chosen for each criterion $i \in N$, such that, for all $x \in X$, $s_i(x)$ indicates the score that criterion i assigns to x . Let S_i be the set of all possible scoring functions for criterion i .

Table 1. Ordinal rankings of Utilitarianism (U), Kantianism (K), and Virtue Ethics (VE) provided by general theory evaluation criteria. If two theories are placed on the same line, an indifference relation holds between the two. Note that this is a purely hypothetical example for illustration.

	General theory evaluation criteria				
	Internal Consistency	Fruitfulness	Scope	Simplicity	Action-guidance
U	U	U	U	U	UK
K		VE	K	K	VE
VE		K	VE	VE	

³Note that the criteria as well as the ordinal rankings have been introduced without any reference to credences. Credences do play, if at all, an indirect role in our procedure. If one finds one (or multiple) strategie(s) compelling for credence ascriptions, it might well be the case that one’s credence goes up if, let us say, consistency of a theory has been established.

⁴The following formal presentation is based on Gaertner and Xu’s (2012) characterisation.

Let P be the set of all orderings over X . A profile $s = (s_1, \dots, s_i)$ is a list of scoring functions, one for each criterion. An aggregation rule f is defined as a mapping: $S_1 \times \dots \times S_i \rightarrow P$. Let $S = S_1 \times \dots \times S_i$.

f is said to be an E -based scoring rule, to be denoted by f_E iff, for any $s \in S$, and any $x, y \in X$, it is the case that

$$x \succsim y \Leftrightarrow \sum_{i \in N} s_i(x) \geq \sum_{i \in N} s_i(y)$$

where $\succsim = f(s)$. The asymmetric and symmetric parts of \succsim will be denoted by \succ and \sim , respectively.

What is important in order to achieve feasibility and comparability of such a grading procedure is that the grades to which the criteria are subjected are general enough. For some criteria a qualitative sequence such as “high, sufficient, still admissible, doubtful, inadequate”, as in our subsequent example, may be fully adequate. In other contexts, the grading structure may have to be more specific, both quantitatively and qualitatively. Also the grading structure has to be based on judgments that can be rationally determined. Furthermore, the evaluative terms need to make sense with regards to every general theory evaluation criterion that is considered. We fully admit that this is a critical step in our proposal, but in real life there is, as we see it, a multitude of examples in which a common grade structure is used to evaluate various objects (remember, for example, the grading structure at school or university or in sports).

To make this general scoring rule more concrete, let us apply it to the previous example. One possible assignment of scores to alternatives and the corresponding resulting ranking could be the following⁵ (Table 2).

Table 2. Rankings of Utilitarianism, Kantianism, and Virtue Ethics given we apply the general scoring rule to the example from Table 1. The numbers in brackets indicate the total scores per moral theory under consideration.

Scores/Criteria	General theory evaluation criteria				
	Internal Consistency	Fruitfulness	Scope	Simplicity	Action-guidance
5 (high)	U			U	U K
4 (sufficient)	K	U	U		
3 (still admissible)		VE	K	K	
2 (doubtful)	VE	K	VE		VE
1 (inadequate)				VE	

$$U(23) \succ K(17) \succ VE(10)$$

⁵All criteria are weighted equally in this general scoring rule. However, the general scoring rule is flexible enough to give particular criteria more weight. This can be done either by dividing criteria up into sub-criteria or by adding weighting parameters which allow calculating weighted scores for every criterion. Let us point out that we do not see “action-guidance” as a killer criterion in the sense that this criterion has to meet a certain threshold, since otherwise an agent might be left without a clear indication of how to act. It is a permissible outcome of our procedure that the agent suspends judgment on what actions should be performed.

We should take the possibility into account that errors occur when evaluating the ethical theories according to the given criteria. These errors might occur since, for example, some of the criteria involve expectations (e.g. fruitfulness) and/or other criteria are not specified clearly enough and, hence, allow some room for interpretation. We can take this possibility of error into account by not only picking the top-ranked theory but a wider set of theories in this first step of our procedure. In the example discussed in **Table 1** and **Table 2**, it looks adequate to dismiss Virtue Ethics and to focus on Utilitarianism and Kantianism. This move can be made more precise, for example, by introducing a minimal requirement or threshold defined in relation to the rank scoring system that is used to evaluate the competing theories (i.e. a minimal aggregated rank score that a theory needs to reach). Alternatively, if a wide difference in aggregate rank scores manifests itself between several of the theories and the rest, there would be good reasons to exclude the latter from further consideration. The advantage of leaving more than one object in the choice set at stage one is that these elements still have an opportunity to prove their adequateness or usefulness at a second stage.

The key question that has to be answered in relation to our procedure is the following: Why does it make sense to transform the ordinal rankings of the criteria, which are used to evaluate the moral theories, into a cardinal ranking and to add up these scores?

This problem was discussed in depth by [Gaertner and Wüthrich \(2016\)](#) in the context of theory choice. We believe that the justification strategy developed there can be applied to our current situation. The basic idea is the following. When agents are asked to evaluate and compare a certain number of competing actions or objects, they start with an ordinal ranking in terms of “better, equally good, and worse” which, at least in some cases, may be reached fairly quickly. By their very nature, these ordinal rankings are relatively coarse. The cardinal scale which is then introduced forces the evaluating person to think more deeply and more accurately about the various characteristics of each competing object, and how each fares in relation to the others. Those who are against any type of cardinalisation may view this mental procedure as arbitrary, others may call this a process of “purification and refinement”. Notice that this process does not change the base information that is given. The cardinality that is introduced leads to further differentiation. It can be interpreted as a medium that introduces a language of qualitative verdicts. The score “5” in **Table 2** may, for example, correspond to the verdict “high”; the score “1” to the verdict “inadequate”. These verdicts in turn can be motivated with respect to the pragmatic projects which we try to achieve when we engage in acting. Accordingly, note that we do not want to say that the moral theories *themselves* provide the resources for these cardinal and comparable scores. Rather, it is the moral theory plus additional information from the concrete decision problem at hand (we call this the *problem context*) that allow moving to a cardinal representation of

the situation. For example, a moral theory can be “sufficiently simple” and “sufficiently fruitful” for solving a particular problem of evaluating alternative actions.

3.2. Step 2: Selecting the Guiding Moral Theory to Assess Different Actions

Given the output of the first step, we have to differentiate between two possible scenarios. Either the set of *relevant* moral theories is a singleton or it contains at least two members.⁶ If the set is a singleton, we no longer face the problem of moral uncertainty. The problem boils down to identifying the ranking of actions in light of this *single* moral theory. If the set is not a singleton, then the problem of moral uncertainty is still given; it has just been reduced in terms of the *amount of theories* that have to be considered.

We suggest proceeding as before starting from an ordinal ranking of theories, which order the actions respectively, under consideration. We again apply the proposed general scoring rule from above to this ordinal information. We can rely on the justification that was given above in section 3.1 to move from ordinal rankings to a cardinal description of the problem.

To make the present step more concrete, let us apply it to Anna’s decision problem. According to the result documented in **Table 2**, Utilitarianism and Kantianism remain in the relevant choice set. These theories are now exposed to additional criteria such as the two in **Table 3**, namely “an equal treatment of human beings” and “integrity and inviolability of the human body”. We call them content criteria. Again, the underlying ordinal rankings are converted into cardinal scores. Let us suppose that both Kantianism and Utilitarianism get a high score in terms of equal treatment but that Kantianism scores higher in terms of inviolability so that we obtain a case of weak vector dominance in favour of Kantianism. Spelling out the particular scores and adding them is therefore not necessary in this case but if one insisted in doing so, **Table 3** reveals that in terms of equal treatment Kantianism and Utilitarianism are on equal par, while in terms of inviolability, the difference in favour of Kantianism is four ranks, which shows, at least in the situation considered, the relative strength of Kantianism in relation to Utilitarianism. Anna should therefore *not* push the fat man from the bridge. One might have added an additional criterion or property, namely the notion of an abstention from any intrusive act on the part of the deciding person, an aspect of negative freedom (Berlin, 1969) which would have made the argument in favour of Kantianism even stronger, but the choice situation is already clear enough along the two criteria given.

⁶Note that our procedure in the first step always generates a non-empty choice set. However, it might be possible that all of the alternative theories are deemed as “inadequate” because the aggregate scores are too low. In such a situation, further theories would have to be considered. This in itself is an important upshot of our grading procedure. Alternatively, of course, the general theory evaluation criteria could be modified but we would regard this as an admission of defeat rather than as a solution to the problem.

Table 3. Rankings of Kantianism and Utilitarianism according to content criteria given we apply the proposed scoring rule. The numbers in brackets indicate the total scores per moral theory under consideration.

Scores/criteria	Content criteria	
	inviolability	Equal treatment
5 (high)	K	K U
4 (sufficient)		
3 (still admissible)		
2 (doubtful)		
1 (inadequate)	U	

$$K(10) \succ U(6)$$

An assessment such as the one in **Table 3** is informative and instructive insofar as a confirmation of the existence of (weak) vector dominance would have said nothing about the *strength* of one criterion over another. Gaps between alternative proposals in the evaluation grid are a mirror of argumentative force. The statement that there is a case of vector dominance would be totally silent on this. We consider the transparency that **Table 2** and **Table 3** display as an advantage and perhaps strength of our proposal. At the same time, it offers “a battleground for the boggled mind”. Is it really true, so the reader may ask, that in Anna’s decision situation, Virtue Ethics fares so badly in contrast to Utilitarianism and the Kantian view? To summarize, we believe that the two-step procedure provides a framework for the articulation of reasons for why a particular action has been chosen as choice-worthy in the case of moral uncertainty.

3.3. Steps 1 and 2 Again: An Evaluation of Specific Moral Theories: Alternative Concepts of Distributive Justice

In what follows we would like to present another example of our two-stage procedure. We consider moral theories that concern alternative concepts of justice. We hope to be able to show that the approach we suggest does not only apply in a sequence of general theory evaluation criteria and content criteria but can also be proposed in cases in which theories are evaluated according to particular properties or requirements that are viewed as fundamental as, for example, in questions of distributive justice. There are two differences to our example in sections 3.1-3.2: First, the following example is not motivated by a concrete choice problem (i.e. the trolley problem). Second, the example with different theories of distributive justice relies on two stages consisting of different criteria. The first stage does not contain general theory evaluation criteria but specific properties instead that are deemed relevant in an evaluative setting. We add this example to highlight the flexibility of our approach and to show, as the reader will see shortly, how our proposal can be used to motivate the search for new theory

proposals. Depending on the area of science one is considering, one can imagine that there are some very basic requirements that will be regarded as minimal or indispensable, and that a set of more disputable requirements exists to which the theories under consideration are then exposed at a second stage.⁷

In order to be brief, we confine our analysis to three theories, namely *Rawlsian Maximin* MR (the focus is on the worst-off in society), *Maximax* MM (the focus is on the best-off), and *Utilitarianism*, concepts that have been frequently discussed both in philosophy and economics. As criteria at stage one of our procedure, we choose “unrestricted domain” UD (all logically possible combinations of n-tuples of utility values are admissible in an n-person society), “weak Pareto” WP (if one vector of utility allocations is component-wise judged better than another, the former vector is to be preferred by society), “anonymity” AN (all utility values that are expressed by the members of society are weighted equally), and a “minimal equity” requirement ME (there is at least one situation where the group of utility-wise worst-off is “socially decisive”, i.e., determines the social outcome among alternative options). Maximin satisfies all these criteria while Utilitarianism has to make certain requirements as to the curvature of the individuals’ utility functions in order to achieve minimal equity. Maximax clashes with ME.

A table similar to the earlier ones would look like this: **Table 4**

Following our proposal, Maximax (MM) will henceforth be disregarded since ME clashes with this concept. At the second stage, we introduce the criterion that a trade-off between individual utilities be admissible over a limited interval of utility values only, more concretely within an interval that lies above a required minimum utility level.

If this criterion is chosen (see **Table 5**), Maximin and Utilitarianism are deemed unacceptable. The latter holds because under Utilitarianism, a trade-off between, for example, the utility values of the “poorest” and the “richest” would be possible, if the richer claim to gain more than the poor would lose, thus making the poorest even poorer in a transfer from poor to rich. The Maximin principle does not permit any trade-offs at all, so that it would not be acceptable anyway.

Table 4. Rankings of Maximin (MR), Maximax (MM), and Utilitarianism (U) under the chosen set of criteria of the first stage.

Scores/criteria	Stage 1 criteria			
	UD	WP	AN	ME
3 (fully satisfied)	MR, MM, U	MR, MM, U	MR, MM, U	MR,
2 (satisfied only under special circumstances)				U
1 (not satisfied)				MM

⁷Note that this is a different task than in the previous example. Here, we are explicitly concerned with the evaluation of theories and not the evaluation of some particular acts.

Table 5. Score assignments to theories of distributive justice under the limited trade-off criterion.

Scores/criteria	Stage 2 criteria
	criterion: limited trade-offs
2 (satisfied)	UF
1 (not satisfied)	MR, U

The upshot of this two-stage procedure is that we are left empty-handed. However, in footnote 6 we pointed out that such a result cannot be excluded *a priori*. Furthermore, we argued that in such a case, alternative concepts would have to be considered. *Utilitarianism with a floor* UF (Rawls, 2001: pp. 127-129) is such a candidate. This version of utilitarianism requires that a minimum level of utility be guaranteed to all layers of society with the proviso that utility trade-offs be limited to utility distributions beyond that level so that at least a certain minimum is secured. Utilitarianism with a floor fulfils all stage 1 criteria and by its very nature, it satisfies the introduced criterion of limited trade-offs (see again Table 5). An issue still unsolved would then be to decide where to fix the required minimum level.

We thus obtain a unique choice or verdict. The trade-off limitation has become a cut-off criterion. We believe that the procedure above provides more analytic validity and substance than a statement which claims that among all four theories of distributive justice from above, a person's credence in Utilitarianism with a floor, for example, is highest.

4. Concluding Remarks

At this point we have to ask whether our approach has practical applications. In sections 3.1 and 3.2 we referred to the trolley problem. This is a deep philosophical issue which is intimately related to what is called a "triage" situation. There are many examples of this kind. To whom should a scarce medicine be given? Which patient(s) should receive intensive respiratory care in the current Covid-19 pandemic? Should a hijacked plane which is heading for a stadium filled with thousands of people watching a soccer match be shot down? Should an autonomous driving vehicle run into older foot-passengers or rather younger persons, into men rather than women, if a deadly accident cannot be avoided? Similarly, as discussed in section 3.3, alternative notions of distributive justice are closely related to economic issues such as a fair tax system. Should poor people, for example, be exempted altogether from paying any income taxes and should very rich persons face a highly progressive tax scheme? Should politicians consider any trade-off relations between lower and higher income brackets, an issue which has immediate influence on the degree of cohesion within a society?

In our view, the two-step procedure proposed in this paper tries to provide an answer to the issues just mentioned, though a debatable one admittedly. From a formal point of view, it has several desirable features. First, our proposal is *constructive in nature*. It delivers an answer to the problem of moral uncertainty without relying on credence *distributions* whose genesis and change over time have not yet been well explored in our view.

Second, the generalized scoring rule has *greater flexibility* than the Borda rule. While the Borda rule requires that each and every criterion rank the alternative theories in a linear order, such a high degree of uniformity is not demanded by the method proposed here. We consider this as an advantage since the single criterion has more flexibility to articulate its intensity and strength.

Third, the suggested aggregation procedure has interesting *formal properties*. To start, the procedure is sensitive towards the degree of criteria fulfillment. This is, in social choice theory, denoted as a form of positive responsiveness or positive association in the sense that a unilateral change in the fulfilment of some criterion in favour of x , let us say, should be reflected on the aggregate level in the same and not in the opposite direction. The model also satisfies a property that is sometimes called consistency, at other times reinforcement (Young, 1974, 1975), demanding that if the set of criteria is split up into two parts and a certain theory (or action) wins in both subsets, then this theory (or action) must also win in relation to the complete set of criteria. Finally, the general scoring function satisfies a reformulated version of Arrow's (1951, 1963) Independence of Irrelevant Alternatives for a cardinal context.

Fourth, the hierarchical structure of our aggregation procedure can serve as a *fruitful background* against which further questions about moral uncertainty can be asked. By proposing a sequence of steps, the procedure highlights assumptions and has the potential for further exploration: For example, what modifications do we have to make when we apply theory evaluation criteria designed for empirical theories to the case of moral theories? Which content criteria can we impose on ethical theories?

One may object that our approach makes very demanding assumptions about the intellectual process the agent allegedly goes through. We agree that some of the Kuhnian criteria, "fruitfulness", for example, may not play the most decisive role in the moral agent's psychology. On the other hand, every serious decision maker should be concerned about the issue of internal consistency, and simplicity is an attribute that will make the search for an adequate action much easier. If this is not convincing, we would like to come back to the criteria in step 2 of our argumentation in section 3.2. Equal treatment of persons and the inviolability of the human body are very basic principles with which every moral theory should be confronted. Their violation would not only question the solidity and acceptance of any moral theory but would also provoke a serious confrontation with jurisdiction in democratically elected and administered societies.

Acknowledgements

I am very grateful to Nicolas Wüthrich for our collaboration and many discussions over the last years. I also wish to thank Alex Voorhoeve for his comments on earlier versions of the paper.

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

References

- Arrow, K. J. (1951, 1963). *Social Choice and Individual Value* (2nd ed.) New York, NY: John Wiley.
- Berlin, I. (1969). Two Concepts of Liberty. In I. Berlin (Ed.), *Four Essays on Liberty*. Oxford: Clarendon Press.
- Cuneo, T., & Shafer-Landau, R. (2014). The Moral Fixed Points: New Directions for Moral Nonnaturalism. *Philosophical Studies*, 171, 399-443.
<https://doi.org/10.1007/s11098-013-0277-5>
- Dietrich, F., & List, C. (2017). What Matters and How It Matters: A Choice-Theoretic Representation of Moral Theories. *The Philosophical Review*, 126, 421-479.
<https://doi.org/10.1215/00318108-4173412>
- Gaertner, W., & Wüthrich, N. (2016). Evaluating Competing Theories via a Common Language of Qualitative Verdicts. *Synthese*, 193, 3293-3309.
<https://doi.org/10.1007/s11229-015-0929-4>
- Gaertner, W., & Wüthrich, N. (2020). *Handling Moral Uncertainty: A Structural View*. Manuscript, University of Osnabrück and London School of Economics.
- Gaertner, W., & Xu, Y. S. (2012). A General Scoring Rule. *Mathematical Social Sciences*, 63, 193-196. <https://doi.org/10.1016/j.mathsocsci.2012.01.006>
- Gracely, E. J. (1996). On the Noncomparability of Judgments Made by Different Ethical Theories. *Metaphilosophy*, 27, 327-332.
<https://doi.org/10.1111/j.1467-9973.1996.tb00212.x>
- Gustafsson, J. E., & Torpman, O. (2014). In Defence of My Favourite Theory. *Pacific Philosophical Quarterly*, 95, 159-174. <https://doi.org/10.1111/papq.12022>
- Hansson, S. O., & Hirsch Hadorn, G. (2016). *The Argumentative Turn in Policy Analysis: Reasoning about Uncertainty*. Zürich, Springer.
<https://doi.org/10.1007/978-3-319-30549-3>
- Kuhn, T. S. (1977). Objectivity, Value Judgment, and Theory Choice. In T. S. Kuhn, (Ed.), *The Essential Tension—Selected Studies in Scientific Tradition and Change* (pp. 320-339). Chicago, IL: The University of Chicago Press.
- MacAskill, W. (2014). *Normative Uncertainty*. DPhil Dissertation, Oxford: University of Oxford. <http://www.williammacaskill.com/>
- Rawls, J. (2001). Justice as Fairness. In E. Kelly (Ed.), *A Restatement*. Cambridge: Harvard University Press.
- Sepielli, A. (2013). Moral Uncertainty and the Principles of Equity among Moral Theories 1. *Philosophy and Phenomenological Research*, 86, 580-589.
<https://doi.org/10.1111/j.1933-1592.2011.00554.x>
- Young, P. H. (1974). An Axiomatization of Borda's Rule. *Journal of Economic Theory*, 9,

43-52. [https://doi.org/10.1016/0022-0531\(74\)90073-8](https://doi.org/10.1016/0022-0531(74)90073-8)

Young, P. H. (1975). Social Choice Scoring Functions. *SIAM Journal of Applied Mathematics*, 28, 824-838. <https://doi.org/10.1137/0128067>