Utility, Progress, and Technology

Proceedings of the 15th Conference of the
International Society for Utilitarian Studies

edited by
Michael Schefczyk & Christoph Schmidt-Petri

KIT Scientific Publishing

Cover design by Nico Brähler

Contents

Contents

Panel Discussion

From Utilitarianism to Prioritarianism

An Empathy-Based Internalist Foundation of Welfare Ethics

Christoph Lumer, University of Siena, Italy

Abstract

The article develops an internalist justification of welfare ethics based on empathy. It takes up Hume's and Schopenhauer's internalistic (but not consistently developed) justification approach via empathy, but tries to solve three of their problems: 1. the varying strength of empathy depending on the proximity to the object of empathy, 2. the unclear metaethical foundation, 3. the absence of a quantitative model of empathy strength.

1. As a solution to the first problem, the article proposes to limit the foundation of welfare ethics to certain types of empathy. 2. In response to the second problem, an internalistic metaethical conception of the justification of moral principles is outlined, the result of which is: The moral value of the well-being of persons is identical to the expected extent of (positive and negative) empathy arising from this well-being. 3. The contribution to the solution of the third problem and focus of the article is an empirical model of the (subject's) expected extent of empathy depending on (an object's) well-being. According to this model, the extent of empathy is not proportional to the expected empathy, but follows a concave function and is therefore prioritarian. Accordingly, the article provides a sketch of an internalist justification of prioritarianism.

I   The Search for a Justification of Utilitarianism and a New Proposal - With a Prioritarian Outcome

The justification of utilitarianism is not exactly a success story. Mill's justifications (1998, ch. 4, par. 3-9), for example, are paradigmatic fallacies. Several justifications, in an intuitionistic, question-begging way, already presuppose certain moral principles – Hare (disguised by semanticism) (1981, sects. 1.3; 1.6) and Singer (1993, 11-12; 2011, 91-93; 100-102; 113-14) presuppose a certain form of universalization, Harsanyi (1953) presupposes ignorance of one's own identity (thereby operationalizing impartiality like Rawls) or the Pareto Principle plus the application of Bayesian Rationality to moral decisions (Harsanyi 1955). Still others build on – questionable – rationality-theoretical premises – in particular the equalization of one's own future time slices and the time slices of other persons (Sidgwick 1982, 381-82; 418-19; Parfit 1992, 281-82; 342; 346; Broome 1991, 231-37; 239-40). Most utilitarians do not even give a reason and only rely on their intuitive acceptance of utilitarianism (e.g.

Smart 1973, 3-8). But the research on the rational foundations of utilitarianism also contains unexploited potential, e.g. Hume's reflections.

This article develops a justification of a welfarist moral value function based on empathy, or, in Hume's (1978, 317-19) terminology, on sympathy. Here I will use the terms "*empathy*", "*sympathy*" and "*compassion*" interchangeably and with them mean: an emotion evoked by considering some person's or sentient being's well-being, that leads to the compassionate emotion, which may be negative or positive, according to the object's assumed negative or positive well-being. My justification takes up Hume's (1978, sects. III.2.2; 3.1-3) and Schopenhauer's (1977, §§15-6) internalistic (but not consistently developed and empirically flawed) approaches, but tries to solve three of their problems. The first problem, seen by Hume himself (but not satisfactorily solved), is: Morality formally requires universality and impartiality, while empathy varies with the temporal, spatial, social and personal distance from the object of empathy (1978, 580-82; 603). The second problem is the unclear metaethical basis of Hume's and Schopenhauer's approaches. The third problem, seen by neither of them, is that empathy is not proportional to the well-being of the empathy object:[1] An empirical study I conducted shows that compassion with negative well-being is more intensive than happiness about others' positive well-being.

My proposal for solving the first problem is that, in order to achieve universality and impartiality, which are necessary for the purpose of morality, the moral justification should be based only on certain universalistic forms of empathy: empathy that arises when considering the effects of one's own actions on the well-being of others (and not, for example, the empathy that arises from direct contact with others) (Lumer 1999). Unfortunately, this is only a very weak component of our total empathy but the only one which is subject-universalistic, i.e. leads to interpersonally identical valuations of the same objects (though there will rarely be valuations of the same object by different persons). The problem of the emotion's and therefore also the appertaining motivation's weaknesses may be resolved by taking the empathic emotion only as the signal which informs us about its object's moral value. This signal then has to be amplified by other motives which follow its lead. The most important such amplifiers are socially valid norms (Schopenhauer also suggested this (1977, 257-58)) and our feeling of moral self-worth. In the following I will not deal any further with this problem but will focus on the first and third problems.

---

[1] Hume, instead, seems to presuppose some proportionality between the pleasure of the persons affected and the spectators' sentiments: sympathy for the affected, love and hate for those changing their fate (1978, 591).

The proposal for the solution of the third problem is to study empirically how the degree of other persons' well-being influences our empathy.[2] More precisely: In the following an empirical model is developed, that calculates which extent of empathy (i.e. the integral of positive and negative empathy over time) occurs depending on the average well-being of an object of empathy. The expected extent of empathy is then the hedonistic and internalist moral reason for empathy-optimizing actions; and this empathy is also the basis and source of the internalist morality: The proposal equates the expected extent of empathy – which is identical to its expected hedonic desirability for the empathetic subject – with the moral value of the object's underlying well-being. The most important outcome of the model below is: Because of the greater intensity of negative empathy, the resulting moral value function is not utilitarian (linear function from well-being to moral desirability) – as a Humean may have guessed –, but prioritarian (concave function from well-being to moral desirability). This means the model provides a justification[3] and quantitative specification of prioritarianism.

In the following I will first (II) briefly explain the metaethical basis of the justification developed here and thereby outline my solution of the second, metaethical problem; this is only for understanding the approach, a further justification of this basis is not possible here. (III) Subsequently, I will present the empirical model of expected empathy in order to (IV) draw normative-ethical consequences.

---

[2] I have developed the model set out below in my professorial dissertation from 1992, which, however, was published only in 2000, 2nd edition 2009 (Lumer 2009). This paper is the first English presentation of the model.

[3] If prioritarianism is justified at all, exceptions aside, it is justified only intuitionistically, in particular as a middle way between utilitarianism, which is economic but does not intrinsically care about distributive justice, and maximin or leximin, which cares about distributive justice by giving priority to those who are worst off but in an extremist way. An exception is Hurley's (1989, 360-82) idea to introduce a risk-averse, concave weighting of prospects into a Rawlsian/Harsanyian framework of rational decision under uncertainty about one's identity. The result would be a concave, today we would say: prioritarian, moral value function. But Hurley did not elaborate this idea nor bring it together with the critique of utilitarianism and Rawls' difference principle; she envisioned her idea as something egalitarian – prioritarianism at that time was not yet a theoretical movement.

II   Metaethical Foundations of the Justification of Morals [4]

What is a valid justification of morals at all? Justifications of morals, firstly, contain an *epistemically rational* component: By justifying these morals, one gains insights which distinguish them as something special. Secondly, valid justifications of morality contain a *practical* component: they are to have the consequence that the addressee of the justification adopts the justified morality as his own and, if possible, also acts on this basis.

The simplest and clearest way to bring the epistemic and the practical requirements together is to design moral justifications as arguments for a thesis about the object of justification, i.e. about the moral principle, etc. However, this cannot be *any* thesis; but the justification for this thesis must meet certain conditions. A thesis which fulfils these conditions is the *justification thesis* for moral principles. In this way, the epistemic requirement can be met by the fact that the justification still consists in an argumentatively valid and adequate argument which leads to a justified belief; and the practical and moral requirements can be met by selecting a particular thesis about the object to be justified, i.e. the justification thesis that this object has a certain justificatory quality *F*. I have developed several adequacy conditions for selecting this property *F*:

Adequacy Condition 1 (AC1): Motivation or practical requirement: Moral justification theses about moral principles are motivating in the sense that if a prudent addressee (i.e.: an epistemically and practically rational addressee with certain relevant information) is justifiedly convinced of the justification thesis (i.e. that the moral principle in question is *F*), he is motivated, at least to some extent, to adopt and observe the moral principle.

The motivation requirement is the specifically *practical* component of the conception for justifying moral principles. It makes the justification internalistic.

Adequacy Condition 2 (AC2): The motivating effect's stability with respect to new information: The motivating effect of a justified conviction of a justification thesis is stable with respect to new information, i.e. it is not lost as a consequence of acquiring additional true information.

Stability with respect to new information is the *rational* component of the concept of justifying moral principles. The only thing we can rationalize (in the sense of making it rational)

---

[4] Unfortunately, for reasons of space, this section is rather apodictic. A detailed explanation and justification of the presented metaethical approach can be found in: Lumer 2009, 30-127; 577-632; 2015.

directly are beliefs, indirectly also actions and other things. And the two main directions of that rationalization are: first, to make our beliefs true, i.e. to acquire possibly only true beliefs (or to correct false beliefs) by observing epistemological rules and, second, to increase the number of true beliefs. The requirement of the motivation's stability with respect to new information introduces the practically relevant maximum of epistemic rationality into the conception of practical justification.

Adequacy condition 3 (AC3): Moral instrumentality: Principles for which the justification thesis is true fulfill the function of moral principles, they meet the instrumental requirements for such principles and for morals in general.

Moral instrumentality is the specifically *moral* component of the conception of justification. If the "justified" moral principles do not fulfill the function of morality we are no longer dealing with a justification of a morality.

What is the function of morality? One can facilitate answering this question by distinguishing the structural components of morality. Normative morality consists mainly of a moral desirability function and moral norms, institutions and virtues. Once the moral desirability function has been established, it can be used to justify the other components of morality as more or less good means for realizing moral values. So, proceeding in this way, only the practical function of the moral value order has to be determined.

One can distinguish an *individual* and a *socially binding morality*, where the latter is designed to regulate social relations in an intersubjectively binding way. Here I will mainly deal with the second type. The sense of a socially binding moral desirability function could be *prudential-consensualistic*:

1. *Consensualistic requirement:* Socially binding moral evaluation criteria constitute a common moral value system that provides the intersubjectively shared standard (i) for assessing socially relevant measures, (ii) for planning social projects and (iii) for consensual arbitration of interpersonal conflicts of interest. In addition, for individuals the purpose or sense of such an intersubjectively shared value system could be to procure a benchmark for self-transcendent ego ideals and actions. I call this quality of the desired moral value functions "*subject universalism*", i.e. the value of all value objects (or more precisely the value relation of every two value objects p and q (= $U(p)/U(q)$) of this value function is roughly identical for all (or nearly all – except e.g. for psychopaths) moral subjects of the moral community.

2. *Prudential requirement:* The prudential requirement is that the subjective value functions to be compared according to subject universalism be parts or components of the subjects' prudential desirability functions. *Prudential desirability functions* express what is good for

the respective subject and hence, rationally or from a prudential point of view, should be the guideline of the subject's decision. Prudential desirability functions are constructed similarly to the utility functions of rational decision theory but with much stricter, philosophically developed standards, which also permit the criticism and correction of the subject's present instrumental or even intrinsic preferences (cf. e.g. Brandt 1979, part I; Lumer 2009, 241-428; 521-48). Prudential desirability functions are intersubjectively different – that I have a headache is mainly bad for me and neutral for you, and the reverse holds for your headache –; otherwise they could not express the *personal* good. Therefore, the subject-universalistic requirement is not intended to refer to complete prudential desirability functions but only to components thereof, i.e. parts of the total value which derives from particular types of consequences of the value object.

This concludes the metaethical considerations regarding the justification of morals; now the exposed conception has to be applied. The next step is empirical, viz. to enquire empirically, with the help of empirical decision theory and moral psychology, which component V of the prudential desirability function U is subject-universal and hence can be adopted as the moral desirability function. The result of a respective scrutiny is that interpersonally (nearly) identical components of our prudential desirability functions arise from our expected *compassion* and our expected *feelings of respect*. Of these two subject-universal feelings and motives, however, compassion is much better suited as the basis of the moral desirability function. For unlike compassion, one can hardly specifically optimize one's feelings of respect; respect is rather passive, it evokes motives for defending the respected object, but not motives for creating or improving respected objects. Therefore, in the following I will develop a model of a prudential desirability function based on empathy, or more precisely: a model of expected empathy depending on the well-being of other people. This expected empathy, in turn, corresponds to its hedonic desirability for the empathic subject. Ultimately, desirability procured through empathy is the sought-after subject-universal component of the prudential desirability function, which defines moral desirability. In short: The extent of expected empathy (according to the empirical model) is equated with moral desirability.

# III An Empirical Model of the Expected Extent of Empathy

So the present task is to develop a – simplified – quantitative model of how the well-being of other persons whom we neither particularly like nor dislike is reflected in our expected

sympathy, i.e. the expected amount of our feelings of positive and negative sympathy. In short the model informs about the (expected) extent of our sympathy depending on other persons' well-being. The model's most important simplifying assumptions are these: 1. The object of our sympathy is the assumed well-being of the person(s) for whom we feel sympathy. 2. Errors in our assumptions about other persons' well-being statistically offset each other. 3. The model deals with *universal* sympathy only, i.e. a kind of sympathy we feel for strangers whom we neither like nor dislike in a particular way and whose behavior we do not judge in a moral way. 4. In a very flexible society like ours, the chances to be confronted with the lot of other people are equal for all objects of sympathy. And the salience of the fate of other people is equally distributed statistically. 5. The intensity of our compassion depends on the intensity and duration of considering it. But again, the expected values of these two quantities are intersubjectively equal for all objects of sympathy. 6. Prudent subjects have feelings of sympathy and do not try to avoid them.

The first step in developing this model is to determine the intensity of our sympathy depending on the assumed condition of the object. Consider figure 1.
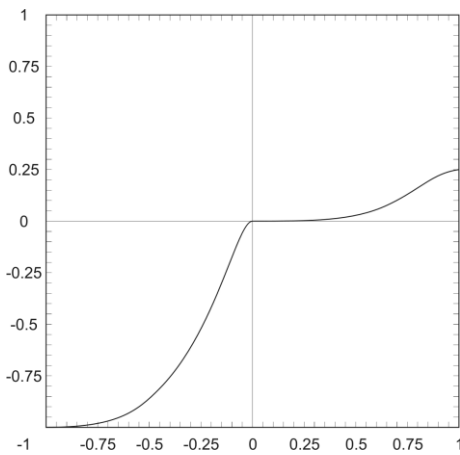
Fig. 1: Sympathy $S(x)$ depending on assumed well-being x
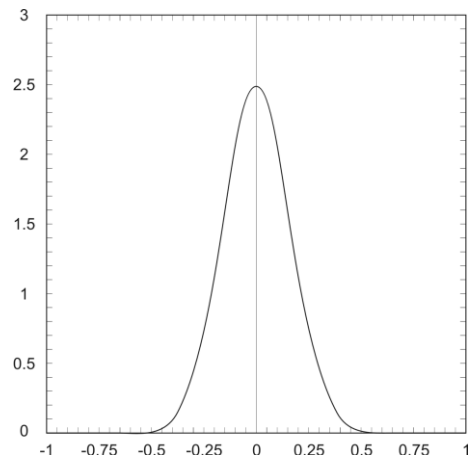
Fig. 2: Distribution $PD(x)$ of well-being x for $x_\mu=0$

The x-axis represents the object's well-being; positive values represent pleasant feelings, negative values represent unpleasant feelings. The y-axis represents the appertaining sympathy, negative values representing pity and positive values representing pleasant feelings of sharing joy or the other person's flourishing. The other person's well-being as well as the sympathy are normalized into the interval [-1;1] with 0 being the point of indifference. Plausible assumptions about the function from well-being to sympathy are: The sympathy function ascends monotonously. To neutral well-being we are sympathetically indifferent; i.e. the function includes the point (0;0). Negative sympathy, i.e. pity, is much more intense

than positive sympathy. At the time when I developed this model I conducted interviews for testing the willingness to exchange packages of such feelings with different durations. This kind of willingness was then hedonistically reinterpreted as the subject's comparative judgement of the respective extents of sympathy. According to these calculations, pity for the most extreme sort of suffering was 4 to 10 times more intensive than positive sympathy with the most extreme form of joy. Conservatively I have taken 4 to be the right relation. The most extreme points of the function of figure 1 then are (-1; -1) and (1;0.25). Empirically our normal well-being ranges between 0 and 0.4; our sympathetic reaction to this kind of normal well-being is minimal. Outside of this region of normalcy sympathy's intensity increases rapidly, though much more rapidly versus negative than versus positive. When approaching extreme states of well-being sympathy will be satiated. – From these assumptions one gets the sympathy function designed in figure 1.

The most important feature of this function is that it is not linear: Pity is much more intense than positive sympathy; and normal states of well-being (between 0 and 0.4) are nearly neglected by our sympathy.

The second step of the model is to find out the intrasubjective distribution of well-being for different objects of sympathy over their life-time. For establishing the extent of sympathy, we need not know the exact course of the object's well-being but only the proportional duration of the single levels of well-being during the whole life. Again simplifying, I assume that these well-being levels are distributed normally. The open parameters of such a normal distribution are, first, the mean μ and, second, the spread σ. Empirical research on well-being has revealed that the intersubjectively most extreme long-term means of well-being of the unhappiest and the happiest people, positively-linearly transformed in our scale (-1;1), lie between 0 and 0.4 ($0 \leq \mu \leq 0.4$), so that the happiest people in the long run arrive at a mean of 0.4. Continuing the simplification, I assume that the mean levels of well-being of happy and unhappy people are intersubjectively different, but that the spread remains the same. Relying on some plausible assumptions about the absolute duration of very extreme feelings, the spread can be calculated as being equal to $\sigma=0.16$. The resulting curve for $\mu=0$ is shown in figure 2. In this way one gets a bundle of curves of normal distributions each representing the distribution of different well-being levels for typical more or less happy individuals; all these curves are equally shaped but their means range from 0 to 0.4 – according to the individual happiness –; i.e. the curves are shifted to the left or to the right with the top of the curves ranging between 0 and 0.4.

The third step is to multiply the probabilities given by the normal distribution of well-being with the sympathy function and to calculate the integral from -1 to 1 over this product function. The result of this operation is the expected extent of sympathy, i.e. the sum of all

feelings of sympathy which one expects to feel for a given person depending on the mean well-being µ of this person. This operation can be repeated for all the long-term means µ of well-being from the empirically expected range of such means, i.e. the interval from 0 to 0.4. The result is the function of the extent of sympathy depending on the long-term mean level µ of well-being. Normalizing the mean levels of well-being as well as the resulting extents of sympathy by a positive-linear transformation into the interval [0;1] one gets the normalized function of the extent of sympathy: ESN(m). This function is represented in figure 3.
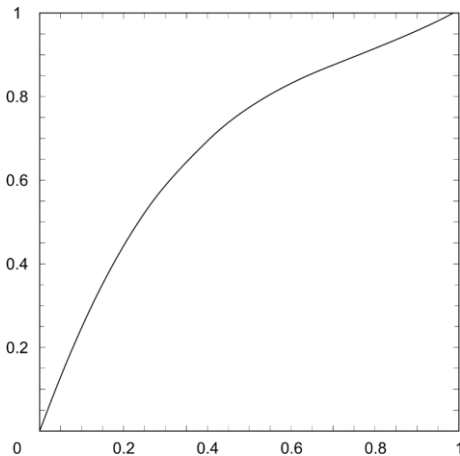
Fig. 3: Normalized extent of sympathy *ESN*(*u*) depending on the long-term mean level *u* of well-being
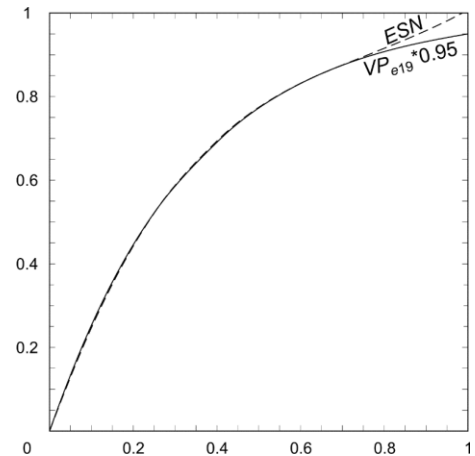
Fig. 4: Comparison of the normalized extent of sympathy *ESN*(*u*) with utilex *VP*$_{e19}$·0.95 (exponential value function)

In this function the *x*-axis represents the normalized lifetime mean-levels of well-being; and the *y*-axis represents the normalized expected extent of sympathy resulting from facing persons having the respective mean-level of well-being.

If somebody wants to value some social order from a purely sympathetic perspective he can assess the various mean levels of well-being of the people living in this society, find out the appertaining extent of sympathy and, finally, sum up these extents of sympathy. This, of course, is the same procedure which a hedonist prioritarian has to use to assess the prioritarian value of this social order. The only difference is that the prioritarian uses the prioritarian welfare function instead of the function of the extent of sympathy.

For formal mathematical reasons, but above all for metaethical reasons, one would like to have functions with certain properties as prioritarian weighting functions: They should be concave throughout, i.e. have a constantly decreasing gradient, rise monotonously, etc. For

this purpose I have discussed several mathematical curve families (Lumer 2005, sect. 3.1). The most suitable of these curve families are exponential curves:

for $e > 1$: $VP_{ee}(u) = e/(e-1)) \cdot (1-e^{-u})$; and

for $e = 1$: $VP_{e1}(u)=u$; this is identical to the right-hand limes of $VP_{ee}(u)$ for $e \rightarrow 1$ (see figure 5).

$VP_{ee}(u)$ is the family of **e**xponential **P**rioritarian **V**alue functions with the parameter e, where "e" within the function is a parameter equal to or larger than 1 (and does not mean Euler's number), which expresses the degree of priorioritarianism: the higher the number e, the stronger the prioritarian inclination. With $e=1$ the prioritarian inclination does not exist; the curve coincides with utilitarianism. With extremely high values for e the function creates leximin preferences. e-values between these extremes represent more or less radical forms of prioritarianism.
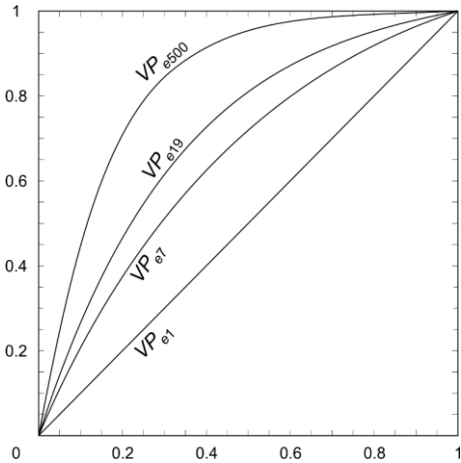

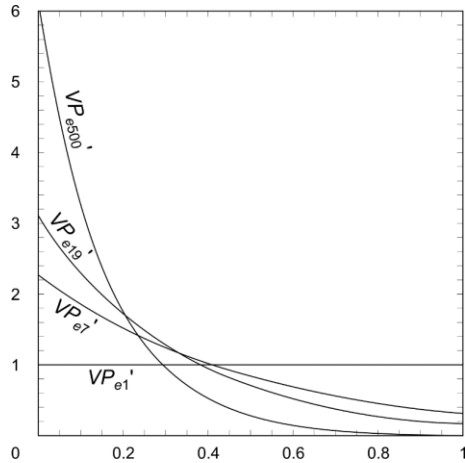Fig. 5a: Exponential functions: VP_{e1}, VP_{e7}, VP_{e19}, VP_{e500} Fig. 5b: First derivations VP_{ee}' of exponential functions

One can now compare the empirically established function of the extent of empathy with these ideal prioritarian curves. The one that fits best is the curve for $e=19$. The two curves are compared in Figure 4. (The prioritarian function has been compressed by the factor 0.95 in order to facilitate the comparison.) One can easily see that, for a big stretch the two functions are more or less identical. That is why I have proposed the exponential prioritarian curve with $e=19$ ($VP_{e19}(u)$) as the internalistically justified prioritarian weighting function.

The function of the extent of sympathy just presented is based on some rather provisional measurements. But its general prioritarian shape is rather stable with respect to changes of

these assumptions and measurement results. So the exact function may be changed by re-measuring but the prioritarian shape will remain, because it depends only on the stronger intensity of pity as compared to positive sympathy.

IV   Conclusion

On the basis of all these considerations we can now draw the conclusion: The internalist justification strategy for value ethics based on the adequacy conditions presented in section 2 and the prudential-consequentialistic determination of the function of socially binding morals, via an empirical scrutiny of possible subject-universal components of the prudential desirability functions has led to identifying empathy with others whom we neither like nor dislike in a particular manner as the sought source of the moral desirability function. On the basis of prudential hedonism, the empirical model of the expected extent of sympathy depending on other persons' (mean life-time) well-being provides the quantitative specification of this prudential desirability function. This function is mathematically simplified as $VP_{e19}(u)$, so that this function is therefore proposed here as the internalistically justified moral value function. This value function is universalistic, welfaristic and prioritarian. In the next parts of the theory, on the basis of this value function, certain moral norms, institutions, virtues, etc. can be justified as good means of realizing moral values.

What has been achieved with the study presented here? 1) If one tries to justify welfare ethics internalistically in the manner outlined above through compassion, the result is a version of *prioritarianism*, not utilitarianism (i.e. a concave not a linear moral value function). 2) In this way, prioritarianism has been justified internalistically, i.e. with recourse to (pre-moral) motives. This goes far beyond a merely intuitive acceptance of prioritarianism. 3) Prioritarianism has been quantitatively specified, beyond a vague comparative intuition, in a way that is needed for complex moral assessments with the comparison of many different consequences for different persons. From an infinite spectrum of more or less radical forms of prioritarianism, a specific one is distinguished as internalistically justified.

References

[1] Brandt, Richard B. 1979. *A Theory of the Good and the Right*. Oxford: Clarendon.

[2] Broome, John. 1991. *Weighing Goods: Equality, Uncertainty and Time*. Oxford/Cambridge, MA.: Blackwell.

[3] Hare, Richard M. 1981. *Moral Thinking: Its Levels, Method, and Point*. Oxford: Clarendon.

[4] Harsanyi, John C. 1953. "Cardinal Utility in Welfare Economics and in the Theory of Risk-Taking." *Journal of Political Economy* 61: 434-35.

[5] _____. 1955. "Cardinal Welfare, Individualistic Ethics, and Interpersonal Comparisons of Utility." *Journal of Political Economy* 63: 309-21.

[6] Hume, David. 1739-40/1978. *A Treatise of Human Nature*. Edited, with an Analytical Index by L. A. Selby-Bigge. 2nd edition with text revised and variant readings by P. H. Nidditch. Oxford: Clarendon

[7] Hurley, Susan L. 1989. *Natural Reasons: Personality and Polity*. New York/Oxford: Oxford University Press.

[8] Lumer, Christoph. 1999. "Intergenerationelle Gerechtigkeit: Eine Herausforderung für den ethischen Universalismus und die moralische Motivation." In *Was heißt Gerechtigkeit?: Ethische Perspektiven zu Erziehung, Politik und Religion*, edited by Reinhold Mokrosch and Arnim Regenbogen, 82-95. Donauwörth: Auer.

[9] _____. 2000/2009. *Rationaler Altruismus: Eine prudentielle Theorie der Rationalität und des Altruismus*. 2nd supplemented edition. Paderborn: mentis.

[10] _____. 2005. "Prioritarian Welfare Functions – An Elaboration and Justification." *Working paper*, Department of Philosophy and Social Sciences, University of Siena. http://www.lumer.info/wp-content /uploads/2020/07/A066_Lumer_PrioritarianWelfareFunctions.pdf.

[11] _____. 2015. "Ethical arguments for moral principles." In *Proceedings of the 8th International Conference of the International Society for the Study of Argumentation, Amsterdam, July 1-4, 2014*, edited by Bart Garssen, David Godden, Gordon Mitchell and A. Francisca Snoeck Henkemans. Amsterdam: Sic Sat. http://rozenbergquarterly.com/issa-proceedings-2014-ethical-arguments-for-moral-principles/?print=pdf.

[12] Mill, J. S. 1861/1998. *Utilitarianism*. Edited by Roger Crisp. Oxford/ New York: Oxford University Press.

[13] Parfit, Derek. 1984/1992. *Reasons and Persons*. Oxford: Clarendon.

[14] Schopenhauer, Arthur. 1840/1977. "Preisschrift über die Grundlage der Moral." In *Werke in zehn Bänden. Zürcher Ausgabe, Vol. VI,* 143-315. Zürich: Diogenes [Engl. Transl.: *On the Basis of Morality*. 2000. Translated by E. F. J. Payne. 2nd edition. Indianapolis: Hackett.]

[15] Sidgwick, Henry. 1874/1982. *The Methods of Ethics*. Indianapolis/ Cambridge: Hackett.

[16] Singer, Peter. 1979/1993. *Practical Ethics*. 2nd [enlarged] edition. Cambridge: Cambridge University Press.

[17] _____. 1981/2011. *The Expanding Circle: Ethics, Evolution, and Moral Progress*. 2nd edition. Princeton/Oxford: Princeton University Press.

[18] Smart, J. J. C. 1961/1973. "An outline of a system of utilitarian ethics." In *Utilitarianism for and against*, edited by Idem and Bernard Williams. 3-74. Cambridge: Cambridge University Press.