Supererogation

Edited by Christopher Cowley

Supererogation

EDITED BY

Christopher Cowley

Contents

Notes on contributors

KEITH ANSELL-PEARSON

Keith Ansell-Pearson is Prof of Philosophy at the University of Warwick and author and editor of books on Nietzsche, Bergson, and Deleuze. His book on Nietzsche and Epicurus will be published by Bloomsbury Press in 2015 and he is currently at work on a study of Nietzsche's reception of Jean-Marie Guyau.

ALFRED ARCHER

Alfred Archer is assistant professor at the University of Tilburg, the Netherlands. His doctoral thesis examined the nature of supererogation and the implications that accepting the existence of supererogation has for a range of different debates in moral philosophy. He has published articles in *Philosophical Studies, Ratio, Philosophia* and *The Journal of Ethics and Social Philosophy*. For up-to-date information about his research visit: http://alfredarcher.weebly.com/

MATTHIAS BRINKMANN

Matthias Brinkmann is a DPhil student at the University of Oxford, from where he also received a BPhil. He works on a thesis on political legitimacy, and is interested in political philosophy more generally, normative ethics, and Kant.

CHRISTOPHER COWLEY

Christopher Cowley is lecturer in philosophy at University College Dublin, Ireland. He is the author of *Medical Ethics: Ordinary Concepts, Ordinary Lives* (Palgrave 2007) and *Moral Responsibility* (Routledge 2013). He is also interested in the philosophy of criminal law.

ELIZABETH DRUMMOND YOUNG

On completing her doctorate at the University of Edinburgh in 2004, Elizabeth Drummond Young taught moral philosophy and history of philosophy there for a number of years. She

Notes on contributors

now teaches courses on the philosophy of love and friendship in the Office of Lifelong Learning of the University and is independently engaged in research on the philosophical and religious aspects of exemplary love. She is interested in both analytic and continental strands of thought. Her most recent publication is a chapter 'Love Reveals Persons as Irreplaceable' discussing the work of Jean-Luc Marion and Raimond Gaita in *Love and its Objects* ed. Christian Maurer, Tony Milligan and Kamila Pacovska (Palgrave Macmillan 2014)

MICHAEL FERRY

Michael Ferry is an Associate Professor of Philosophy at Spring Hill College, USA. He works in moral and political philosophy with a particular interest in how we define the limits of duty and how we justify options in the space of moral and practical reasons. His previous work on supererogation has been published in *Philosophical Studies*.

CHRISTOPHER HAMILTON

Christopher Hamilton is Lecturer in Philosophy of Religion in the Department of Theology and Religious Studies at King's College London. He is the author of four books, including *Living Philosophy: Reflections on Life, Meaning and Morality* (Edinburgh University Press, 2001) and *Middle Age* (Acumen, 2009), and has published articles on Kierkegaard, Nietzsche, Simone Weil, Primo Levi, and in aesthetics, ethics and philosophy of religion.

DAVID HEYD

David Heyd is Chaim Perelman Professor of Philosophy at the Hebrew University of Jerusalem. He has written extensively on supererogation, including *Supererogation* (CUP 1982), the Stanford Encyclopedia entry on the subject and a few articles on supererogation in bioethics, on its relation to toleration and forgiveness, and on supererogatory promises.

D. K. LEVY

D. K. Levy is a philosophy teaching fellow at the University of Edinburgh where he works on moral philosophy. His research concerns understanding, especially moral understanding.

Areas in which he has published include Wittgenstein, moral philosophy, and cognitive psychology.

ROWLAND STOUT

Rowland Stout is Associate Professor at UCD School of Philosophy, having previously been at the Universities of Oxford and Manchester. He works on the philosophy of mind and action, epistemology, metaphysics and ethics. He is the author of *Things That Happen Because They Should* (OUP, 1996), *Action* (Acumen, 2005) and *The Inner Life of a Rational Agent: In Defence of Philosophical Behaviourism* (Edinburgh University Press, 2006). He is currently editing a book entitled *Process, Perception and Action.*

ULLA WESSELS

Ulla Wessels is Professor of Practical Philosophy at the Philosophy Department of the Saarland University in Saarbruecken. She works in several areas of practical philosophy, mainly in ethics. Her publications include *Das Gute* (Frankfurt: Klostermann, 2011), as well as *Die Gute Samariterin: Zur Struktur Der Supererogation* (De Gruyter 2003).

MARK WYNN

Mark Wynn is Professor of Philosophy and Religion in the School of Philosophy, Religion and History of Science at the University of Leeds, where he is also Director of the Centre for Philosophy of Religion. He is the author of *Renewing the Senses: A Study of the Philosophy and Theology of the Spiritual Life* (Oxford University Press, 2013), *Faith and Place: An Essay in Embodied Religious Epistemology* (Oxford University Press, 2009) and *Emotional Experience and Religious Understanding: Integrating Perception, Conception and Feeling* (Cambridge University Press, 2005). His current interests include the distinctive goods that are the object of spiritual practices, the epistemic significance of the notion of tradition, and the relationship between the different vocabularies that are used to map the various phases of the spiritual life.

Beyond the Call of Duty: The Structure of a Moral Region

ULLA WESSELS

A woman risks her life to save someone else's child from a house that is on fire. While in his prime, a man donates one of his kidneys to a dialysis patient whom he does not know. In Auschwitz, Maximilian Kolbe sacrifices his life for the life of another prisoner.

We tend to call these actions good or even very good, but not obligatory. We tend to say that the people performing them are going beyond the call of duty – that what they do is supererogatory.

Supererogation originally attracted attention within Catholic ethics. Remember the biblical story of the Good Samaritan, who came across a man who had been beaten up und robbed. He bandaged the man's wounds, took him to an inn, and giving the innkeeper two silver coins he said: 'Take care of him, and whatever more you spend, I will repay you when I come back.'[1] Although the expression 'whatever more you spend' (in Latin: 'quodcumque supererogaveris') only refers to the expenses of the innkeeper, the Church Fathers agreed that in essence the story is about the efforts of the Good Samaritan. By giving the innkeeper two silver coins and declaring his willingness to reimburse him for any extra expenses, the Good Samaritan did more good than could be asked of him. According to the Church Fathers, he was following not only the Ten Commandments, but also what they called 'consilia', the divine instructions to saints, which go beyond the Ten Commandments.[2]

From Catholic ethics, supererogation spilled over into secular ethics, initially living in the shadows, though. Only towards the end of the 19th century it experienced a renaissance: within Alexius Meinong's value theory[3] and later on in the writings of J. O. Urmson, Roderick M. Chisholm, David Heyd, Gregory

---

[1] Luke 10:35.

[2] See, e.g., Augustine, *De sancta virginitate* (in *Patrologiae cursus completus*, ed. by J.-M. Migne, vol. 40, Paris: Garnier, 1841; *De sancta virginitate* written about 401), chap. 5.

[3] Alexius Meinong, *Psychologisch-ethische Untersuchung zur Wert-Theorie* (Graz: Leuschner & Lubensky, 1894).

Mellema, Paul McNamara, and many others.[4] Those authors were less concerned with the question precisely which real-life actions are supererogatory[5] than with theoretical matters: what is the *structure* of a morality that features supererogation,[6] and what is the *structure* of supererogation itself?

Views on the structure of supererogation have become more and more sophisticated over the years. However, the majority of them is still built around a model, which I propose to call the 'threshold model for supererogation'. According to the threshold model, there is an amount of good we are obliged to do, and by doing something that is better, we are going beyond the call of duty. In other words:

> In every situation, there is a threshold for the good to be done such that, firstly, it is obligatory to perform an action that meets the threshold, and, secondly, every action that exceeds the threshold is supererogatory.[7]

[4] J. O. Urmson, 'Saints and Heroes', in *Essays on Moral Philosophy*, ed. by A. I. Melden (Seattle: University of Washington Press, 1958); Roderick M. Chisholm, 'Supererogation and Offence', *Ratio* 5 (1963), 1–14; David Heyd, *Supererogation* (Cambridge: Cambridge University Press, 1983); Gregory Mellema, *Supererogation, Obligation, and Offence* (New York: State University of New York Press, 1991); Paul McNamara, 'Making Room for Going Beyond the Call of Duty', *Mind* 105 (1996), 415–450.

[5] Nor with the other question the Fathers of the Church had on their mind: how to use the surplus of merits the saints gained by their supererogatory actions? As is generally known the answer to this question led to the selling of indulgences.

[6] This question has been asked for all major types of ethics: for consequentialism or even utilitarianism, e.g., by Douglas W. Portmore, 'Position-Relative Consequentialism, Agent-Centered Options, and Supererogation', *Ethics* 113 (2003), 302–332, and Jean-Paul Vessel, 'Supererogation for Utilitarianism', *American Philosophical Quarterly* 47 (2010), 299–319; for Kantian ethics, e.g., by Jens Timmermann, 'Good but Not Required?', *Journal of Moral Philosophy* 2 (2005), 9–27, and Marcia Baron, 'The Supererogatory and Kant's Wide Duties', in *Reason, Value, and Respect*, ed. by Robert Johnson and Mark Timmons (Oxford: Oxford University Press, 2015); and for virtue ethics, e.g., by Gregory Mellema, 'Moral Ideals and Virtue Ethics', *Journal of Ethics* 14 (2010), 173–180, and Claire Michelle Brown, *Supererogation for a Virtue Ethicist*, PhD dissertation (Notre Dame: University of Notre Dame, 2011).

[7] It is most clearly presented and defended by Michael Slote (*Common-Sense Morality and Consequentialism* (London: Routledge & Kegan, 1958), chap. 3, and 'Rational Dilemmas and Rational Supererogation', *Philosophical Topics* 14 (1986), 59–76) and James Dreier ('Why Ethical Satisficing Makes Sense and Rational Satisficing Doesn't', in *Satisficing and*

At first sight, the threshold model looks plausible enough. Yet it is wrong, as I will try to show in sections 1 and 2 of this paper. In section 3, I will present an alternative: a model that is, as I hope, right or, at any rate, less wrong than the threshold model.[8]

1. Dismissing the Threshold Model

The threshold model is wrong because it overlooks an important fact: some actions are impermissible and thus do not deserve to be called supererogatory even though they are better than some that do.

Let us start with an example, a purely fictitious one so that we will not be torn between normative and empirical questions. Imagine, firstly, that you want to donate some money to save lives, and the more money you donate, the more lives you will save. Imagine, secondly, that there is a largest sum that is the minimum you are morally required to donate – the threshold sum –, and there are larger sums that exceed the threshold sum. Imagine, thirdly, that within the realm of these larger sums there are two that behave as follows: one is *slightly* larger than the other is, but you will save *far* more lives by donating this slightly larger sum.

Since the example is still a little abstract, I will now flesh it out, call it situation S, and illustrate it in Figure 1. In situation S, if you don't donate anything, you won't save a single life; if you donate €50, you will save one life; if you donate €5,000, you will save 100 lives; if you donate €10,000, you will save 101 lives; and if you donate €10,050, you will save 200 lives.

---

[8] Throughout this paper I use the expressions 'going beyond the call of duty' and 'doing something that is supererogatory' as if they were synonymous – although Paul McNamara ('Supererogation, Inside and Out', in *Oxford Studies in Normative Ethics*, vol. 1, ed. by Mark Timmons (Oxford: Oxford University Press, 2011)) has recently argued that they are not. If McNamara is right about this, the model I will suggest is a model for 'going beyond the call of duty' rather than for supererogation. For example, McNamara deems supererogation (unlike going beyond the call of duty) to depend on the agent's intentions in a certain way that is not covered by that model. I believe that the sensitivity to intentions that McNamara has in mind could be incorporated into it, but that, of course, is a different matter.

possible actions	donations in €	numbers of lives saved
f_5	10,050	200
f_4	10,000	101
f_3	5,000	100
f_2	50	1
f_1	0	0

Figure 1. Situation S

Now imagine, fourthly, that if you donate one of the larger sums (€5,000, €10,000, or €10,050), it doesn't make much difference to you whether you donate another €50 or not. And imagine, finally, that you are obliged to donate at least €50, and that donating €5,000 or more is doing more good than can be asked of you. In Figure 1 the dotted line indicates this.

Under those conditions, I submit, if you donate exactly €10,000, you will be doing something that is *not* supererogatory. Granted, you are already doing more good than can be asked of you. Nevertheless, you will be doing something that is *impermissible* because with merely €50 extra you could save 200 lives instead of 101. And since supererogatory actions are not impermissible,[9] you will be doing something that is *not* supererogatory.

In other words, you do not always have a free choice in the realm of actions by which you do more good than can be asked of you, for this realm still features *conditional obligations*. It is required that, *if* you are already going to donate €10,000 or more – which in itself is not required! –, then you throw in an additional €50, given that the additional €50 are negligible to you whereas the additional moral benefit is huge. Should you fail to meet this conditional obligation, you fall into what I propose to call a *supererogation hole*. The metaphor of a hole seems appropriate because in such cases a non-supererogatory action is 'surrounded', so to speak, by supererogatory

[9] Supererogatory actions are good or even very good, but not obligatory. In doing them the agent goes beyond the call of duty, and to say that the agent goes beyond the call of duty, but does something that is impermissible would be odd. After all, if the agent goes beyond the call of duty, she at least fulfils all the obligations that apply – not matter what else she does. Hence, whatever supererogatory actions may be, they are *permissible*. This has been pointed out by many, e.g. by David Heyd, op. cit. in note 4, 120–5, and by Paul McNamara, op. cit. in note 8, 219.

actions. Although you are moving in a *neighborhood* of actions by which you would exceed the obligatory, you risk performing one that fails to even deliver the obligatory.

What situation *S* shows, then, is, firstly, that supererogation depends on the interaction of at least two factors: what is at stake morally, and what is at stake for the agent;[10] and, secondly, that this interaction is more subtle than the threshold model can allow. To repeat, according to the threshold model, there is an amount of good we are required to do, and everything beyond that amount is supererogatory. However, as we have seen, that is not always so. In a sense, a strenuous action can be *both* better than required and too bad to be permissible.

2. Objections and Replies

The threshold model should be dismissed – that is the provisional diagnosis. In order to confirm the diagnosis, I will discuss three objections that could be raised against it.[11] First, though, I will get out of the way a proto-objection.

The proto-objection takes issue with the choice of quantities in situation *S*: the numbers of euros donated and of lives saved, it says, come nowhere near a plausible illustration of the deontic claims. The answer is that this may well be so. However, those who

---

[10] Some think otherwise, among them Michael Slote (op. cit. in note 7, 47f): 'A satisficing concern for good results [...] [is] permissible in some cases [...] where there is no issue of personal sacrifice on the part of the agent'. However, most authors agree that supererogation also depends on what is at stake for the agent – even if they disagree *how* much has to be at stake. While David Hyde (op. cit. in note 4, 2), e.g., thinks that sometimes even small favours suffice, M. W. Jackson ('The Nature of Supererogation', *The Journal of Value Inquiry* 20 (1986), 289–296) and many others believe that much greater efforts like 'repeated sacrifice of self-interest' or even risk of losing one's life are required. Saying that the size of the agent's sacrifice matters does not commit those authors to departing from the threshold model, though. The threshold is one of goodness (when enough good is done, every further grain of good done is supererogatory), but nothing in the threshold model excludes (i) that the threshold differs from situation to situation and (ii) that the question where the threshold lies in a situation depends on the costs the agent would incur by performing the various actions open to her.

[11] For further objections see Ulla Wessels, *Die gute Samariterin* (Berlin: de Gruyter, 2002), chap. 1.5.2.

Ulla Wessels

believe it is obligatory and *not* supererogatory to donate €5,000 to save 100 lives should reduce the number of lives saved; and those who believe it is not obligatory to even donate €50 to save a life should reduce the donations. The point is the *structure* – in particular, that in the range of prima facie supererogatory actions one action may mean a major loss for the agent and, compared to some other actions, a small gain for morality, whereas another action may mean a small loss for the agent but, compared to some other actions, a large gain for morality. As long as this structure is preserved, everybody is welcome to alter the quantities as she or he sees fit.

The three genuine objections cover a wide range. While the first two deny that you will be doing something you should not do if you donate exactly €10,000, the last one admits this, but pleads for sticking to the threshold model (or some variety of it) nevertheless.

The first objection runs as follows. Obviously, the donation of exactly €10,000 *is* supererogatory. For if even the donation of exactly €5,000 is supererogatory, how could the donation of exactly €10,000 fail to be? After all, it saves one more life.

She who believes in the existence of supererogation holes replies: The donation of exactly €10,000 saves one more life indeed and is therefore *better* than the donation of exactly €5,000. That much is uncontroversial. Yet the donation of €10,000 is not supererogatory because it does not meet *all* the obligations that apply in situation *S*. It is required that if you're going to donate €10,000 or more, then you donate *more*, namely at least €10,050. The fact that an action violates an obligation, and an obligation that is – although conditional – final, suffices to make it impermissible and hence not supererogatory. It is one thing for an action to be better than an action that is supererogatory, and another thing for it to be supererogatory itself.

The second objection disputes that the donation of exactly €10,000 is *impermissible*. After all, if it were impermissible it would fall into the same deontic category as the action that is worse than the donation of exactly €50, and this seems implausible. Somebody who donates €10,000 instead of €10,050 should not be tarred with the same brush as somebody who refuses to donate anything at all.

She who believes in the existence of supererogation holes disagrees, and here is why. If I pinch my brother for no reason at all and against his will, I am doing something that is impermissible, and if I kill 100 people for no reason at all and against their will, I am also doing something that is impermissible. Yet it is far worse to kill 100 people than to pinch my brother.

The problem is – if it really is a problem – that classificatory terms are blind to certain quantitative differences. Take for instance the

92

classificatory term 'rich'. It can be used for both a run-of-the-mill millionaire and Bill Gates, with no implication about the difference in their wealth. Something similar applies to refusing to donate anything at all and donating exactly €10,000. The classificatory term 'impermissible' can be used for both actions, with no implication at all, neither a positive nor a negative one, about the difference in their value.

Of course, the classificatory term 'impermissible' could be extended into a comparative one so that refusing to donate anything at all would be impermissible *to a higher degree* than the donation of exactly €10,000.[12] Yet even then the donation of exactly €10,000 wouldn't stop being impermissible just because refusing to donate anything at all was impermissible to a higher degree – just as the run-of-the-mill millionaire doesn't stop being rich just because Bill Gates is even richer.

The third objection warns us not to get carried away. It says: If we are serious about conditional obligations, they can accumulate so that no limits whatsoever apply to the burden imposed on the agent – the whole point of supererogation is undermined.

She who believes in the existence of supererogation holes replies that this objection is serious indeed. How serious, can be seen from a variation of situation S, which I call situation S^*. In situation S^*, beyond the donation of €10,050, each additional €50 save another 100 lives – see Figure 2.

About situation S, we said: it is required that if you are already going to donate €10,000 or more you throw in an additional €50 and donate €10,050. In situation S^* this way of looking at things could be repeated over and over, and so in the end we would have to conclude: it is required that if you are already going to donate €10,000 or more you donate a *very* large amount – say, €50,000 or more. Though each extra €50 may only be small change, these amounts of €50 will soon add up and may well reach a sum you can ill afford. Do we really want to say that if you are prepared to make a certain sacrifice beyond the call of duty, you are obliged to embark on a life of hardship – maybe through to the bitter end? And if so, why resort to the concept of supererogation in the first

---

[12]   Something along these lines has recently been suggested by Martin Petersen (*The Dimensions of Consequentialism* (Cambridge: Cambridge University Press, 2013)) and by others before him – e.g., by Björn Eriksson ('Utilitarianism for Sinners', *American Philosophical Quarterly* 34 (1997), 213–228).

possible actions	donation in €	number of lives saved
f_n	50,000	80,100
⋮	⋮	⋮
f_6	10,100	300
f_5	10,050	200
f_4	10,000	101
f_3	5,000	100
f_2	50	1
f_1	0	0

Figure 2. Situation S^*

place? After all, it is the very job of supererogation to protect us from morality demanding too much of us.

We appear to face a dilemma. *Either* we withdraw the claim that in the original situation S donating exactly €10,000 is impermissible (because a lot more good can be done with just an additional €50) and therefore not supererogatory. *Or* we stick to that claim and proceed to say that in situation S^* not only the donation of exactly €10,000, but also the donations of exactly €10,050 etc. are all impermissible (because in *each* case a lot more good can be done with just an additional €50), and that there is a duty to donate €50,000. Both these options appear unattractive.

We might be tempted by a simple solution here. The simple solution consists in accepting that there is *a situation-independent limit to the agent's burden*. While the threshold for the moral gains may differ from situation to situation, depending on the agent's burden, there comes a point at which a person is so badly off or has already done so much for morality that there is no effort, however small, that can be additionally demanded of her for a moral gain, however large, and that point is the same for all situations.

Yet this simple solution has a major disadvantage. The modal space is large, and dreadful things can happen in it. Irrespectively of how badly off somebody is or how much she has already invested, a moral gain is always conceivable (e.g., preventing the death of many) for which we can justifiably impose another duty on her (e.g., to undergo yet another moment of pain). However, this would not be allowed by a theory of supererogation that accepted a situation-independent limit to the agent's burden. Just as a *result* is simply not '*good*' enough' if far better results can be achieved with

little additional effort, a *sacrifice* is simply not '*large* enough' if far better results can be achieved with little additional effort. The proposed theory underestimates how complex the relation can be between the moral quality of the results on the one hand and the individual sacrifices on the other. Sometimes a simple solution is too simple.

Yet what other solution is available? – Supererogation boils down to granting the agent, as an agent, a *discount* in matters of morality. Yet we need not stop at a discount. We could also give the agent a *bulk* discount, an additional discount for cases in which she is already burdened. Hence, we could ascribe to the agent for the *same* moral yield fewer additional obligations the worse off she is or the more she has already done for morality. In situation S^*, this could, for example, have the consequence that on the one hand you are obliged to add another €50 if you're already going to donate €10,000 or more – while on the other hand you are obliged to add merely another €25 if you're already going to donate, let's say, €11,000 or more. In many areas of life, there is nothing wrong with a bulk discount, and why shouldn't ethics be one such area? Although with a bulk discount the conditional obligations may continue to accumulate and even exceed any limit for the agent's burden, they will do so more slowly. Seen thus, a bulk discount *is* one way out of the dilemma – and, it seems to me, the only acceptable one.

Those were my replies to three genuine objections that could be raised against the provisional diagnosis that we had reached at the end of section 1. The three are, I think, the strongest objections, and so by rebutting them we can regard the provisional diagnosis as confirmed: the threshold model for supererogation is mistaken because it denies the existence of supererogation holes.

3. Presenting an Alternative to the Threshold Model: The Format

Let me now introduce an alternative to the threshold model. I call it the Format. As the name suggests, the Format is not a complete theory of supererogation, but merely a framework for such theories. Since it is more complicated than the threshold model, I will develop it in two steps.

In step 1, I will present the Format for Two, which is concerned with situations that involve only two actions. In situations with more than two actions, various complications arise, and the

Ulla Wessels

possibility of supererogation holes is one of them. In order to tackle those complications we need to deploy the machinery that is at the core of the Format for Two several times in several ways. I will explain this in step 2.

3.1 When only two actions are involved: The Format for Two

We start by considering situations that contain only two actions, which we call f_i and f_j, and ask: is one of these actions supererogatory in such a situation? The Format for Two provides the shape of the answer:

$\exists$ threshold z $\in$ $\mathfrak{R}$
$\exists$ function SM: $\mathfrak{R}^4 \rightarrow \mathfrak{R}$
$\forall$ agents a
$\forall$ actions f_i and f_j:

(F2.1) SM monotonically decreases in the 1st and 4th argument
(F2.2) SM monotonically increases in the 2nd and 3rd argument
(F2.3) $\forall$ $(x_1, x_2, x_3, x_4) \in \mathfrak{R}$: SM $(x_1,$
$x_2, x_3, x_4) < z$ if $x_1 - x_2 < 0$ or $x_3 - x_4 \leq 0$
(F2.4) super$_{\{fj, fi\}}(f_j)$ iff: SM $(u(f_j),$
$u(f_i), u_a(f_i), u_a(f_j)) > z$

While the quantifiers pose not much of a riddle, the four clauses (F2.1) to (F2.4) could do with some explaining. We will best look at them in reverse order.

If we abstract from the leeway that the existential quantifiers leave, we can say that the fourth clause, (F2.4), specifies the necessary and sufficient conditions for action f_j being supererogatory in a situation containing just f_j itself and f_i. The heart of the fourth clause is the function SM. This function takes as its arguments four values and assigns to them a fifth value:

$$\text{SM}\left(u(f_j), u(f_i), u_a(f_i), u_a(f_j)\right) = m.$$

The four input values reflect the two factors on which super-erogation depends: what is at stake morally, and what is at stake for the agent. More precisely speaking, the function SM takes as its arguments the moral values of the actions, $u(f_j)$ and $u(f_i)$, and the values the actions have for the agent, $u_a(f_i)$ and $u_a(f_j)$, for short: their subjective values. The value m that SM assigns to that quadruple is the supererogation measure of f_j relative to f_i

96

– or, as we may put it, the degree to which f_j is supererogatory 'with respect to f_i'.

That way of speaking is, I admit, unusual. Usually we think of an action not as being supererogatory *with respect to another action*, but as being supererogatory *in a situation*, where the situation is an entity that involves one or more further action(s). I do not suggest to change this. I introduce the notion of an action being supererogatory with respect to another action just to give it an auxiliary role, the aim still being to shed light on the notion of an action being supererogatory in a situation. The importance of the auxiliary notion may be less clear in the Format for Two, but will become clearer in the construction of the Format itself.

(F2.4) tells us: Action f_j is supererogatory in a situation that contains just the two actions f_i and f_j if and only if the supererogation measure of f_j with respect to f_i is higher than the threshold z. The threshold z is, unlike the threshold of the threshold model, nothing but a real number. It is the task of a *material* theory of supererogation to specify the supererogation measure and the threshold so that they capture how the interplay of f_j's and f_i's moral and subjective values bears on the supererogatoriness of f_j 'with respect to f_i'. Our current task is more formal: not to provide that specification, but to show how such a specification can be put to the service of a theory that says when an action is supererogatory in a situation and that allows for the possibility of supererogation holes.

The third clause, (F2.3), expresses a constraint on the supererogation measure SM and the threshold z. We can call it the threshold clause. Suppose that action f_j were morally worse than its alternative f_i. In that case, it would not be supererogatory, even if it were very strenuous for the agent. An action is supererogatory, compared to an alternative, only if at least nothing is morally lost by it. And similarly, suppose that action f_j were as least as attractive for the agent than its alternative f_i. In that case, it would not be supererogatory either, even if it were morally much better. An action is supererogatory, compared to an alternative, only if it imposes a burden on the agent.[13]

---

[13] Is the threshold condition plausible? One might be inclined to think that it is at the same time too weak and too strong – that, on the one hand, an action cannot be supererogatory unless something is morally *gained* by it and that, on the other hand, an action can also be supererogatory if it doesn't impose a burden on the agent. Think, however, of a situation in which a woman risks her life to save someone else's child from a house that is on fire – she does so instead of the owner who would have done the same so that nothing is morally gained by her action. Her action is supererogatory

The threshold clause makes sure that action f_j is *not* supererogatory with respect to its alternative f_i if f_j is morally worse than *or* at least as attractive as f_i for the agent; and it does so by making sure that the supererogation measure of action f_j is *below* the supererogation threshold if the difference between the moral values of f_j and f_i is smaller than zero *or* the difference between the subjective values of f_i and f_j is smaller than or equal to zero.

The first two clauses, (F2.1) and (F2.2), are requirements on the function SM. Suppose that the condition in the third clause is met: compared to the alternative f_i, action f_j is morally at least as good and more strenuous for the agent. Then we are the more inclined to call f_j supererogatory the *less* is at stake morally and the *more* is at stake for the agent; we are the more inclined to call f_j supererogatory the *smaller* the difference is between the moral values of f_j and f_i and the *larger* the difference is between the subjective values of f_i and f_j. In other words, think of two situations that differ with respect to the moral gain the agent can achieve with the same effort. We are more inclined to talk of supererogation in the situation in which the moral gain is smaller – achieving the larger moral gain with the same effort may well be just the agent's duty. And now think of two situations that differ with respect to the effort the agent has to make for the same moral gain. We are more inclined to talk of super-erogation in the situation in which the effort is greater – making the smaller effort for the same moral gain may well be just the agent's duty.

So this is the Format for Two. It claims something about the shape of the answer to the question whether an action is supererogatory in a situation that contains only that action and one alternative. It says that there are a function SM and a threshold z with certain properties, including the property that one of the two actions is supererogatory if the function, fed with the appropriate pieces of information about the two actions, exceeds the threshold. The Format for Two may not be much of a revelation in its own right, but it is a crucial stepping-stone on the way to the Format.

---

all the same. And, similarly, think of a situation in which a man donates one of his kidneys to a dialysis patient whom he does not know – because he has three of them and would be better off with just two. Since his action does not impose a burden on him, it is not supererogatory.

3.2 When more than two actions are involved: The Format

Next, we generalize – from situations involving only two actions to situations with more than two actions, f_1 to f_n. We assume, without loss of generality, that if one action has a higher index than another it is at least as good as the other. And again we ask: is one of the actions supererogatory in such a situation? The Format – as did the Format for Two concerning a smaller class of situations – provides the shape of the answer. How the Format goes about it, is illustrated by Figure 3.

My central claim is this: in order to answer the main question – is f_j supererogatory in a situation that contains the actions f_1 to f_n? – we have to ask and answer three sub-questions, each of them employing the notion of one action being supererogatory *with respect to another action* that we have studied earlier (and which for situations involving just two actions needed to be employed only once). The answer to the main question is 'yes' if and only if the answer to each of the three sub-questions is 'yes'. Each of the three sub-questions, that is, identifies a necessary condition for f_j being supererogatory, and jointly the three conditions are sufficient.

Is f_j supererogatory in the following situation?

It is if and only if the answers to three sub-questions are yes.

(1) Is there an action f_i such that f_j is supererogatory with respect to f_i?

(2) Are all the actions that are morally better than f_j supererogatory with respect to f_j?

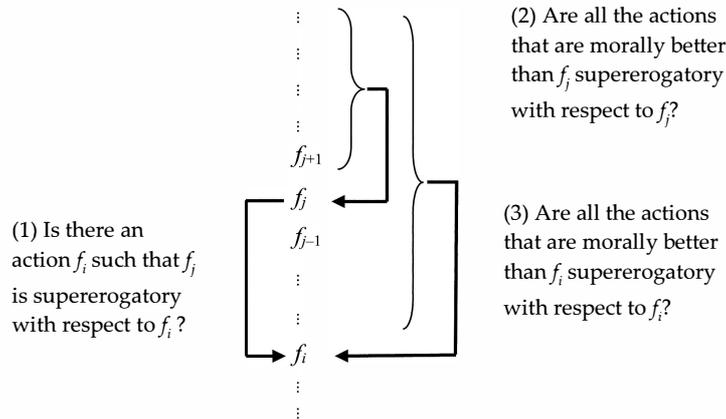(3) Are all the actions that are morally better than f_i supererogatory with respect to f_i?

Figure 3. How the Format provides the shape of the answer to the question whether action f_j is supererogatory in a situation that contains more than two actions

The first sub-question is whether there is an action f_i such that f_j is supererogatory with respect to f_i. If there is no such action, then f_j is definitely *not* supererogatory in such a situation. (Since the action f_i, with respect to which f_j is supererogatory, can at best be as good as f_j, the arrow belonging to the first question in Figure 3 points downwards.)

The second sub-question is whether all actions that are better than f_j are supererogatory with respect to f_j. That second sub-question is the main point of the construction. The first sub-question allowed us to 'look down' at the actions that are worse than (or as good as) f_j and check whether there is at least one among them with respect to which f_j is supererogatory. By contrast, the second sub-question allows us to 'look down' at f_j *itself* and check whether the actions that are better than f_j are all supererogatory with respect to f_j. If not, then there is at least one action that is considerably better than f_j and hardly places any additional burden on the agent. In that case it is not permissible to do f_j instead of something better; it is required that *if* the agent is going to do something that is at least as good as f_j, then she performs one of the actions that are better than f_j. The better actions exercise an 'obligational pull' away from f_j.[14] Action f_j seemed supererogatory, but is not.

While the obligational pull alone does not entail that a supererogation *hole* gapes at f_j, it creates that possibility. A supererogation hole

[14] Here are two real-life examples for such an 'obligational pull': a person who works twenty hours a week for a charity that spends a considerable share of the donations it receives on its own administration – though she could change her mind and work the same number of hours for another, more efficient organization; or a company that provides a certain amount of money to fund a youth centre while knowing that the amount does not quite suffice to also fund a position for a social worker who would make an immense difference to the good that the youth centre can achieve.

Of course, the examples are bound to raise new questions. Perhaps the company deliberately donates an amount that does not quite suffice just in order to make the local council realize its own responsibility. And perhaps the person who spends twenty hours a week working for an 'inefficient' charity does so because this is the only charity towards which she feels a strong affinity. Such things occur. But it also occurs that, along the lines illustrated by the examples, the fact that someone is going to perform an action at least as good as a certain prima facie supererogatory action generates an obligation to further the good even more than she would by performing *that* action.

gapes at f_j only if, although f_j is not supererogatory, an action worse than f_j is. This can be the case when both actions are very good, but the worse one is not subject to such a pull, whereas the better one is. It is easy to prove that the Format allows for that possibility – an action worse than another action can be supererogatory while the other action is not.

Now, one might think that, since there is no conditional obligation to perform an action that is better than f_j, we have already reached our goal. After all, it seems that f_j is thereby permissible and, in view of the answer to the first sub-question, also supererogatory. So why do we need a third sub-question?

We need a third sub-question because the same kind of constellation that endangers the permissibility of f_j also endangers the permissibility of the yardstick f_i. Just as f_j may be impermissible because there exists an alternative to f_j that is far better and hardly places any additional burden on the agent, f_i may also be impermissible because there is an alternative to f_i that is far better and hardly places any additional burden on the agent. But if action f_i is impermissible, then clearly it cannot bear the weight placed upon it by the first sub-question – namely the weight of being the action that 'makes' f_j supererogatory. Action f_i can only bear this weight if it is permissible, and so the third sub-question is: are all actions that are better than f_i also supererogatory with respect to f_i? If not – that is, if among the actions that are better than f_i there is at least one that is not supererogatory with respect to f_i – we are required to perform an action that is better than f_i, and f_j, even if supererogatory with respect to f_i, may well be, ultimately, not supererogatory.

In a nutshell, action f_j is supererogatory in the situation under consideration if and only if the answers to all three sub-questions are yes: firstly, there is an action f_i with respect to which f_j is supererogatory; secondly, all actions that are better than f_j are also supererogatory with respect to f_j; and thirdly, all actions that are better than the yardstick f_i are also supererogatory with respect to f_i.

I hasten to add that, while it was useful for expository purposes to list and discuss the three sub-questions separately and in that order, it was logically not quite hygienic. The real home of the third sub-question is the scope of the existential quantifier in the first sub-question. The first and the third sub-question should be read as the one long question whether there is an action f_i such that (a) f_j is supererogatory with respect to f_i and (b) all actions that are better than f_i are supererogatory with respect to f_i.

At this point, we can assemble our thoughts into the Format:

$\exists$ threshold $z \in \mathfrak{R}$
$\exists$ function SM: $\mathfrak{R}^4 \to \mathfrak{R}$
  $\forall$ agents a
  $\forall$ actions $f_1, \ldots, f_n$
  $\forall$ $f_j \in \{f_1, \ldots, f_n\}$:

(F.1)   SM monotonically decreases in the 1st and 4th argument

(F.2)   SM monotonically increases in the 2nd and 3rd argument

(F.3) $\forall$ $(x_1, x_2, x_3, x_4) \in \mathfrak{R}$:
      SM $(x_1, x_2, x_3, x_4) < z$ if $x_1 - x_2 < 0$ or $x_3 - x_4 \leq 0$

(F.4) $\text{super}_{\{f_1, \ldots, fn\}}(f_j)$ iff:
     (F.4.1) $\exists f_i \in \{f_1, \ldots, f_n\}$:
           $[\text{SM}(u(f_j), u(f_i), u_a(f_i), u_a(f_j)) > z$
           and $\forall$ $f_k \in \{f_1, \ldots, f_n\}$:
           $u(f_k) > u(f_i) \to \text{SM}(u(f_k), u(f_i), u_a(f_i), u_a(f_k)) > z]$
    and

     (F.4.2) $\forall$ $f_k \in \{f_1, \ldots, f_n\}$:
           $u(f_k) > u(f_j) \to \text{SM}(u(f_k), u(f_j), u_a(f_j), u_a(f_k)) > z.$

The first three clauses are old acquaintances: they are identical to the first three clauses of the Format for Two, with (F.1) and (F.2) as the requirements on the function SM, and (F.3) as the threshold clause.

(F.4) is about the core issue, the supererogatoriness of f_j in the situation. (F.4) is a conjunction, the first conjunct of which contains another conjunction. Within (F.4.1), the first conjunct represents the positive answer to the first of our three sub-questions: there is an action f_i such that f_j is supererogatory with respect to f_i. The second conjunct represents the positive answer to the third sub-question: all actions that are morally better than f_i are supererogatory with respect to f_i. Thus, action f_i – the yardstick for f_j – is itself permissible, and so another condition for f_j being supererogatory is satisfied.

(F.4.2) represents the positive answer to the third sub-question: all actions that are morally better than f_j are also supererogatory with respect to f_j. We thus know that f_j is permissible, and only with that information, too, in place are we entitled to say, finally, that f_j is supererogatory in the situation.

So action f_j is supererogatory in a situation that contains the actions f_1 to f_n if and only if, firstly, there is an action f_i such that f_j is supererogatory with respect to f_i (the SM-value of f_j with respect to f_i is higher than the threshold z), and all actions that are morally better than f_i are supererogatory with respect to f_i (their SM-values with

possible actions	donations in €	numbers of lives saved
f_4	10,000	101
f_3	5,000	100
f_2	50	1
f_1	0	0

Figure 4. Situation S'

respect to f_i is higher than the threshold z); and, secondly, all actions that are morally better than f_j are supererogatory with respect to f_j (their SM-values with respect to f_j are higher than the threshold z). This is what the Format says – the format for theories of supererogation that, unlike the threshold model, does justice to the fact that there are supererogation holes.

One of the features of the Format that deserve to be mentioned is this: the Format allows that some actions that are not supererogatory in one set of actions are supererogatory in another set that is a subset of the first one. Consider situation S' in Figure 4.

Situation S' has one action less than situation S – action f_5 is missing –, and while the donation of €10,000 is not supererogatory in situation S, it *is* supererogatory in situation S', because in situation S' there is no alternative to the donation of €10,000 that saves far more lives without being considerably more burdensome for the agent.

The fact that the Format allows this is not a bug; it is a feature of the situations and some basic intuitions the Format is compatible with. In situation S as well as in situation S', the donation of €10,000 is better than the donation of €5,000 because it saves one more life. At the same time, it is impermissible (and thus not supererogatory) only in situation S because in situation S, other than in situation S', a certain further action is available: the donating of merely €50 extra which saves *far* more lives without being considerably more burdensome. So the Format allows what it should allow.

4. Conclusion

Let us look back. We started off with two claims – firstly, that there are actions that do not deserve to be called supererogatory even though they are better than some that do; and secondly, that the threshold model, because it excludes this possibility, should be dismissed. By using an example, we then illustrated and defended the

first claim against three objections. The second claim follows from the threshold model and the first claim. We then set out to develop an alternative to the threshold model that takes into account the existence of supererogation holes: the Format.

Is the Format the last word on supererogation? Unfortunately not. Firstly, the Format is not a complete theory of supererogation, but merely a format for theories of supererogation. The step from the Format to a complete theory of supererogation would consist in opting for a specific function SM and a specific threshold z. Secondly, the Format entails that many things are irrelevant (at least for the step from the evaluation of actions to their deontic assessment) that, when supererogation is concerned, we might not want to be irrelevant – for example, the agent's intentions or the distinction between acts and omissions.

So not only is the Format just a format. Even as a format, it might still not be the last word. But it is a better word than the threshold model, and worth pursuing.

Universität des Saarlandes·Germany
ulla@wessels.de