The Weight of Self-Love in Benevolence and Virtue

HANDELN MIT BEDEUTUNG UND HANDELN MIT GEWALT

Philosophische Aufsätze für Georg Meggle

Herausgegeben von Christoph Fehige | Christoph Lumer | Ulla Wessels

mentis

Fehige/Lumer/Wessels (Hrsg.)
Handeln mit Bedeutung und Handeln mit Gewalt

Einbandfoto von Georg Meggle, Damaskus 2008.

Handeln mit Bedeutung
und Handeln mit Gewalt

Philosophische Aufsätze
für Georg Meggle

Herausgegeben von
Christoph Fehige
Christoph Lumer
Ulla Wessels

Inhalt

IV   Terror, Völkermord und Intervention

V   Leben und werten

VI   Zeit und Raum

The Weight of Self-Love in Benevolence and Virtue

CHRISTOPH FEHIGE

This train of thought is dedicated to Georg Meggle, who has long acted from concern for others, far and near, not least from concern for me. Let us suppose that one day Georg finds himself wondering, as I suspect benevolent and virtuous people sometimes do, whether it might not be a good idea to begin showing some more concern for himself. If he began doing so, would this make him less benevolent, less virtuous? I want to explore some of the theoretical background of this question.

The exploration does not extend to the notorious question whether everything that looks like concern for others is, deep down, concern for oneself. I will assume that this is not so. Both concerns exist, and the challenge is to specify whether and how different degrees of concern for oneself register in benevolence or virtue. Bishop Butler, Francis Hutcheson, and others have shed much light on that nexus, but there is more to say.

1. Self-Including and Self-Excluding Benevolence

A person can be concerned for her own good – which, for the purposes of this paper, may be identified with her own happiness or her own welfare. Philosophers tend to call that concern for her own good the person's self-love. Used in this way, the term »self-love« is not evaluative; whether the presence or absence of self-love or of certain amounts of self-love is good or bad are open questions. If we follow that terminology and start with the weight of self-love not in virtue but in benevolence, we get a fairly clear question to begin with: Does an increase in a person's self-love amount to an increase in her benevolence, or to a decrease, or to neither the one nor the other?

How could an increase in self-love qualify as an increase in benevolence? As far as I can see, it could do so only if self-love were part of benevolence, so that we should ask whether that condition is fulfilled. The answer is less obvious than we might hope.

On the one hand, we could understand benevolence as the desire that people be happy. We could then argue that, for each individual, your desire that the respective individual be happy is one fibre of your benevolence. Your desire that John be happy is one such fibre; your desire than Peter be happy, another; and so forth. But you yourself are one individual, too. So your desire that you yourself be happy, your self-love, is also one fibre of your benevolence.

We tend to reason similarly not in all, but in many other contexts that involve parts or portions or instances. Your love of museum *A*, your love of museum *B*, and so forth, each of these loves amounts to one fibre of your love of museums. In much the same way, every desire of yours that there be rain on a particular day is one fibre of your love of rain, and every preference of yours for a sweet dish, one fibre of your love of sugar. In each of these cases, we consider your relevant general love to be the more substantial the more and the stronger such fibres it is made up from. So why should your desire that people be happy behave differently? If that desire does not behave differently, and if that desire is what we mean by your benevolence, then your self-love is one of the fibres of your benevolence.

On the other hand, in ordinary parlance the word »benevolence« has a strongly altruistic ring. If, for example, as evidence for Mary's benevolence I adduced Mary's tendency to pamper herself, my interlocutors would react with surprise. Such reactions weigh in on the other side – they suggest that benevolence is not the desire that people be happy, but the desire that *other* people be happy, and as such does not comprise self-love.

The first consideration points towards a notion of benevolence that we can call self-including: a person is benevolent to the extent to which she wants the good of others *and* of herself. The second consideration points towards a notion of benevolence as self-excluding: a person is benevolent to the extent to which she wants the good of people *other* than herself. As far as I see, the existence of both kinds of considerations does not pose a problem. We are under no pressure to adjudicate between the self-including and the self-excluding notion. We can cultivate both concepts if at the same time we cultivate an awareness of their difference and a habit of signalling which of the two is being used.

2. *Virtue as Self-Including or Self-Excluding Benevolence*

The terminological flexibility I am recommending in the case of benevolence is a fine thing, provided we do not confuse it with moral flexibility. When the time comes to make moral judgements and when we praise benevolence and rank it among the moral virtues, we had better make clear what it is that we are praising and ranking. At that point the conceptual pros and cons of self-including and self-excluding benevolence that we have encountered in the previous section will reappear as moral pros and cons in the theory of virtue.

The problem presents itself most forcefully if we go so far as to say that benevolence is ultimately the one and only moral virtue; Richard Cumberland and Francis Hutcheson are among the thinkers who have gone that far. It will be easiest if we discuss the matter in such a simple, monistic setting. I will simplify further by abbreviating »moral virtue« to »virtue« without the adjective, hoping that we can bear in mind all the same that the problem has to do with views about the scope and content of morality. Figure 1 shows the question in the virtue-theoretic form in which we now face it. It is, in essence, the question whether self-love is part of virtue.

Let us first look at the positive answer – at the claim that virtue is self-including benevolence. It would follow that, ceteris paribus, Mary would manifest virtue by being good to herself. She would count as more virtuous if she booked herself a massage and as less virtuous if she did not. Should there be deontologists who want to express roughly the same view, they could do so by saying that Mary has a special kind of »duty towards herself«: the moral duty to promote, ceteris paribus, her own welfare. To many of us, this situating of self-interest in the content of virtue or duty seems problematic. It may be warranted to say that the person who shows little or no interest in her own happiness has some kind of deficit, but it sounds odd to say that the deficit is a moral one.

So maybe virtue is self-excluding benevolence, is directed just at the good of others? But that, too, would hurt, although in a different way. It would imply that a person who gives any positive weight at all to her own good displays a lack of virtue. This would be implied because a person who gives some positive weight to her own good would in some dilemmatic situations give priority to a large extra quantity of her own good over a small extra quantity of the good of others, and if virtue were the wanting of the good of others every instance of such a priority

everybody's good

others' good   own good

self-excluding
benevolence   self-love

virtue ?

self-including
benevolence
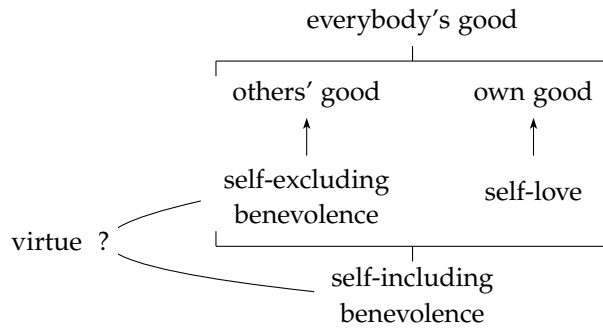
Figure 1: Virtue, two kinds of benevolence, and self-love

would count as a lack of virtue. Suppose, for example, that Mary could secure 1 unit of pleasure for Rose only by undergoing 10,000 units of pain herself, but refuses to do so. Do we really want to say that her decision to avoid the large pain amounts to less virtue than the opposite decision? Is it unvirtuous to take into account one's own pain at all? Again, we hesitate to agree.

We thus have a choice between two conceptions of virtue as benevolence, each with its own potential for irritation. We can say that it is the concern for everybody's welfare that counts (concern for one's own welfare included), and accept, grudgingly, that some regard for one's own welfare increases one's virtue. Alternatively, we can say that it is the concern for other people's welfare that counts, and accept, no less grudgingly, the reverse: that any regard for one's own welfare at all – leading, as it does, to less than absolute priority for other people's welfare – is a dent in one's virtue.

3. Groundwork for a Theory of Virtue as Benevolence

There is another possibility. I will now set forth for a theory of virtue as benevolence that avoids the irritations. I will approach the form and the content of the theory in three stages. This section looks at the idea that underlies the form of the theory. The next section presents that form and illustrates it by combining it with a provisional content, which, in the subsequent section, will be replaced with a content that solves the trouble.

The form of the proposed theory is inspired by the idea that a person's virtue can, in principle, be read off from her preferences concerning specific situations. To use the technical term, the relation is one of supervenience. There is no difference between two people's virtue without a difference in their preferences concerning specific situations.

A situation is, for these purposes, a finite set of at least two possible worlds, with each world having only finitely many inhabitants. A person's preference concerning a situation is her favouring some of the worlds that the situation consists of. For example, Mary's situational preference concerning the four-world situation $\{w_3, w_6, w_{42}, w_{159}\}$ might consist in her favouring, from among those four worlds, worlds 3 and 42. I won't always bother to mention points of time, but should observe the proprieties and do so at least once: situational preferences are preferences at a certain point of time, and so the project is to read off a person's virtue-at-point-of-time-*t* from her situational preferences-at-point-of-time-*t*.

We are not limiting ourselves to situational preferences that, in some way or other, have found their way into the preferrer's consciousness. When it comes to determining a person's virtue, it is difficult to see how the mere fact that a certain existing preference happens not to be on display in the preferrer's consciousness would provide a reason for leaving it out of the picture. We thus include information about preferences that are purely implicit – preferences that do not show up on the person's own radar, but that she has and that would show up on her radar if we made her consider the matter.

Neither are we restricting ourselves to preferences concerning situations that the person runs into, or is likely to run into, in the course of her life. Such a restriction would give us a theory, not of the person's virtue, but of something different: her virtue *in so far as it gets a chance to manifest itself* or in so far as it is *likely* to get a chance to manifest itself. These, too, are interesting topics, but they are not our interesting topic. In order to see how different the topics are, we can imagine an instance: Mary favours, concerning every situation from her real life (or every situation that she is likely to encounter in her real life), the worlds that contain the greatest amount of happiness, but these worlds happen not to involve even the tiniest sacrifice on her part. The restriction at issue would then leave us no way of taking into account whether she is willing to make a sacrifice for the greatest good. This would result in a highly incomplete or misleading picture of Mary's virtue. That Mary desires or does

the best provided it can be had free of charge does not tell us nearly enough.

With these explanations in place, the claim that virtue supervenes on situational preferences should meet with little resistance. The appeal of this claim lies partly in examples of close connections between the two realms. Suppose, for instance, that for every conceivable situation Mary favours the outcome, or one of the outcomes, with the least amount of good. That information about situational preferences would hardly be compatible with the proposition that Mary is virtuous. Or suppose that for every conceivable situation involving worlds with different amounts of good Mary would favour a world with less good than Rose does. That would hardly be compatible with the proposition that Mary is more virtuous than Rose.

Once such examples of connections are acknowledged, it is hard not to go all the way and acknowledge full-fledged supervenience. For if we draw on the preferences concerning every situation (not just situations that the person under consideration finds herself in or is likely to find herself in), then what could we mean by a piece of virtue – more precisely, by a piece of virtue as benevolence – that leaves no trace in those preferences at all? If the person is virtuous in that she wants some good state of affairs to obtain, shouldn't we be able to find a situation in which that state of affairs tips the balance, so that she prefers the world in which the state obtains to one in which it does not? And if she wants that state of affairs more strongly than her neighbour does, shouldn't we be able to find a situation in which that difference in strength tips the balance, so that she prefers a world that the neighbour does not prefer?

These questions, but also general reflections on the nature of attitudes and on the way we talk about them, make it doubtful that a person's virtue and the totality of her situational preferences could come apart. And even if they could, it would remain doubtful that the kind or the part of virtue that does not show up in situational preferences at all is the kind or part of benevolence that we are interested in. It would seem that, to the degree that the attitudes come apart, we are leaving the domain of practical reason, one way or another. We are looking either at virtue that is rationally irrelevant or at a person who is irrational in that her preferences between worlds fail to correspond to her rationally relevant virtue. I will excuse myself from a debate about the status of such oddities and will simply exclude them from the scope of the inquiry. I will study the kind of virtue that is, or virtue in so far as it is, correlated with the person's situational preferences. If there is a kind or portion

of virtue that does not show in the totality of a person's preferences concerning real or unreal situations, it is not the virtue I'm after.

The approach to virtue that has been outlined in this section risks sounding more ambitious than it is meant to be. While an interest in the supervenience of virtue on situational preferences is likely to tempt us to speak of a person's virtue as »made up from« or »composed of« her situational preferences, such phrases should be met with caution. It is not claimed here that preferences between entire worlds take pride of place – that they loom particularly large or are particularly real or basic in the human mind. The opposite may well be true: that »really« or »basically« people care for states of affairs, and that their preferences between worlds are, or should rationally be, informed by the question which states of affairs that they care for hold in the worlds at issue. In that sense, their preferences between worlds may even be informed by, inter alia, their virtue and their benevolence. All these possibilities remain intact. This project utilizes the logical constraints that situational preferences impose on virtue, but there is no commitment to the nature or origin of the situational preferences themselves, nor to the ontological or psychological priorities between the two domains.

4. What the Theory Would Look Like If Virtue Were Self-Excluding Benevolence

We are on our way to a satisfactory theory of virtue as benevolence. Building on the thoughts about supervenience that lie behind us, I will now present what seems to me a suitable form for such a theory by presenting a theory that already has that form, albeit an overly simple content. The makeshift content will identify virtue with one of the two kinds of benevolence distinguished in section 1: with self-excluding benevolence. In the next section, the makeshift content will make way for the ultimate content, which is more complex and more adequate.

We have considered the supervenience of virtue on situational preferences, with an eye to harnessing that supervenience for a theory of virtue as benevolence. The supervenience suggests a two-step procedure. First we come up with a characteristic number for every situational preference that the person whose virtue is at issue harbours; then we connect those specific preferences, via their numbers, to her virtue as a whole. Let us concentrate on the first step and express more precisely what it amounts to. The first step should take us to a function that as-

signs numbers to triples of the form ⟨person, point of time, situation⟩, where the number captures how much good the person wants, at the time, concerning the situation. I propose to call such a number a situational virtue value and such a function a situational virtue function. When I speak of a person's situational virtue function, I am referring to a situational virtue function for that person for some point of time – a function defined for the triples that all involve the person and the point of time, but that each involve a different situation.

We will now grope our way towards such a function. Suppose we show Mary a finite set of at least two possible worlds, all with only finitely many inhabitants, and ask her which world or worlds from that set she favours. Suppose that Mary favours exactly one world. Since we are interested in virtue as benevolence, we should turn our attention to the amount of happiness in that world. But since in this section we are interested in virtue as *self-excluding* benevolence, the amount should exclude Mary's own happiness in that world. We should look at the sum of others' happiness in the world in question. That amount is what we should jot down for future use.

By looking at the favoured amount of others' good without looking at the favoured distribution of the good among those others, the theory of virtue of benevolence will have a utilitarian flavour. But everything that happens in this tract could also be put into the service of a less utilitarian outlook. If we work not simply with the amount of others' good, but with the weighted amount, where the weight is some measure of the equality of the distribution of the good among the individuals, we capture a more egalitarian view of virtue as benevolence. We could also weight each individual's good so that an increase of an individual's good becomes the more significant the worse off that individual is and work with the sum of the weighted individual good, thus capturing a prioritarian view of virtue as benevolence. In this way, the options known from the debate about the moral evaluation of outcomes can be imported into the debate about the moral evaluation of the attitudes towards outcomes. But the utilitarian look at the sheer amount of good that is favoured (or in this case at the sheer amount of others' good that is favoured) is the simplest approach, worth studying first.

Which number shall we jot down if Mary has more than one favourite world in the set, and if these favourites contain different amounts of others' good? For such cases we should make sure that the number reflects all those amounts. We can achieve this by using the average of these amounts. Since we are only looking at sets with a finite number

of elements, Mary cannot favour more than finitely many worlds from that set, and so the average of the sums of others' good in the worlds she favours is well-defined.

We have thus assigned a telling number to each of the specific pieces of information that are waiting to be connected to Mary's virtue in general. We have assigned to each finite set of at least two finitely populated worlds (more precisely, to each triple ⟨Mary, point of time t, situation⟩ that has such a set as its third component) the average of the sums of others' good from Mary's favourite worlds in that set. This mapping captures, for every situation, Mary's all-things-told preference regarding the amount of others' good in that situation. The preferred amount itself serves as the situational virtue value, and the assigning of those values to the situations is the situational virtue function.

A person's virtue, I propose, is measured by the sum of her situational virtue values. If, for example, the sum of Mary's situational virtue values is x, and the sum of Rose's situational virtue values is y, and x is larger than y, then Mary is virtuous to the degree x, Rose virtuous to the degree y, and Mary more virtuous than Rose. Admittedly, the proposal begs for a pinch of salt, since there are infinitely many situations and thus, presumably, infinitely many situational virtue values. Either we find an argument to the effect that each person under consideration has only finitely many situational preferences, which would imply that for each person under consideration we are dealing with only finitely many situational virtue values after all (one such value per situational preference of hers) and that, therefore, summing them up and comparing the sums is no trouble. Or we find out how the spirit of addition and addition-based comparison can be carried into the realm of infinitely many summands – of even more than countably infinitely many summands – and adapt the wording of our proposal to those findings.

Whatever we choose to do in order to tame infinity for the theory of virtue, it will not be done in this paper. The proposal in this paper will treat the structural problems within the situational virtue values themselves. Whether we end up aggregating those values by proving their number to be finite and adding them up straightforwardly or by imitating addition as best we can in the realm of infinite data streams should make no difference. I will assume that we will find some way to aggregate so that the message written into the aggreganda – into the situational virtue values themselves – will not get lost. My way of assuming this is to present the theory, on this occasion, as strictly finitistic: as processing a finite number of situations, each consisting of finitely

many possible worlds that are each inhabited by finitely many sentient beings.

The mathematics of infinity aside, we have now captured a person's self-excluding benevolence. We have done so by consulting every situational preference the person has. It has become clear that every such preference comprises an all-things-told preference for a certain amount of others' good; and that we can understand the totality of these amounts as a measure of the person's self-excluding benevolence and thus, provided that virtue is self-excluding benevolence, as a measure of her virtue.

But now the time has come to remember, from section 2, that it is not all that satisfactory to see virtue as self-excluding. If virtue were self-excluding, giving any weight at all to one's own welfare would amount to a lack of virtue. This would be so because any such weight will have the effect – the reasonable and morally perfectly acceptable effect, it seems – that in some situations a large amount of one's own welfare will take precedence over a tiny amount of other people's welfare, so that for these situations the lesser amount of others' welfare would be preferred. But then virtue, understood as self-excluding, would be lower. Remember the example in which Mary can secure for Rose 1 unit of pleasure only by undergoing 10,000 units of pain herself. Making that sacrifice does not seem better than not making it; it is not even better by utilitarian standards, which are notorious for their demandingness. But if virtue were self-excluding, then, even though making the sacrifice is not better, a person who did not make it would lower her situational virtue value (by 1) and would, to that extent, count as less virtuous.

In short, the picture of virtue as self-excluding remains problematic. We have obtained a focussed version of one view of virtue, but that view is not entirely plausible. We should now look for something that is just as focussed but more plausible.

5. Hybrid Virtue

Hybrid virtue, I suggest, is the solution. The theory of hybrid virtue has the same form as the theory of virtue as self-excluding benevolence presented in the previous section. Like that theory, it exploits the supervenience of virtue on situational preferences by assigning and aggregating situational virtue values. But it has a different content. It works with a different, hybrid characterization of the situational virtue values. The

virtue value of a situation is now defined as the smaller of the following two values:

(1) *favoured sum of others' good* (more precisely, the average sum of others' good in the preferrer's favourite worlds);
(2) *sum of others' good if the sum others' good + own good were maximized* (more precisely, the average sum of others' good in the worlds with the largest sum of good).

We may focus our discussion on cases in which the preferrer favours exactly one world (this concerns value 1) and in which exactly one world contains the largest sum of good (this concerns value 2). The appeals to averages extend things to the other cases along the same lines as were explained in the previous section.

Suppose for a moment, contrary to the facts, that we characterize the situational virtue value not as the minimum of values (1) and (2), but simply as value (1), the favoured sum of others' good. This would amount to equating virtue with self-excluding benevolence. We have seen in section 2 that the equation has much to recommend itself: virtue would reflect concern for others and would not rise with concern for oneself. But we have also seen, towards the end of sections 2 and 4, a weakness. Thus, if we want to understand and appreciate the hybrid characterization of the situational virtue value, we need to check whether it succeeds, by switching from value (1) to value (2) when (2) is smaller, in avoiding the weakness, while keeping the strengths, of the simpler equation.

The weakness that we have noted and want to avoid has to do with the costly way in which Mary could help Rose. The weakness is that a person would qualify as more virtuous by giving less than equal rather than equal weight to her own good. This weakness has two ingredients. The first is the true claim that, if the person gives less than equal weight to her own good, she will, in many dilemmatic situations (situations in which others' good stands in the way of hers), favour a larger sum of others' good than she would if she gave equal weight to her own good. This claim about the favoured sum of others' good is innocent enough on its own, because it does not mention virtue. Only if we combine this claim with the claim that the favoured sum of others' good is situational *virtue*, the larger sum of others' good that is favoured in this way translates into a higher situational virtue value. This translating is what we want to prevent. Whenever the favouring of an extra amount of the good

of others requires that one counts oneself for less than one, we want to prevent that extra good of others from counting as extra virtue.

The way to do this is to cap the situational virtue value. As noted before, the idea of looking at the favoured sum of others' good as that value is on the right track. But the idea needs supplementing with the clause that the value must never exceed that sum of others' good that would be favoured if the maximum sum of good, including the preferrer's own good, were favoured. For any favoured sum of others' good that exceeds that limit, the amount by which it exceeds the limit is bought by discriminating against oneself and should therefore not increase the virtue value. Our dual characterization of the value makes sure that it doesn't. Value (1) makes sure that virtue increases with the favoured amount of others' good, but the switch to value (2) makes sure that this happens only up to a point: the point where one would do more harm to oneself than good to others.

Table 1 shows the theory of hybrid virtue in operation. The example is a situation in which Mary is the agent and Rose the only other party – and in which one of Mary's options is the problematic sacrifice that we have already heard of. There are altogether four options, for each of which Rose's and Mary's welfare are shown, Rose's marked with an »R« and Mary's unmarked below Rose's.

What would be Mary's situational virtue value if she favoured option 1, if she favoured option 2, and so forth? The theory of hybrid virtue claims that, for each such favouring, the situational virtue value is the minimum of values (1) and (2), a minimum that we have already discussed in general terms and that we should now compute for the case at hand. We begin with ingredient (2) of each minimum: with the sum of others' good that would obtain if the largest sum »others' good plus own good« obtained. Column three shows that the largest sum »others' good plus own good« (that is, the largest sum of Rose's and Mary's good) obtains in option 3, and the welfare diagram for option 3 shows that there the sum of others' good (that is to say, Rose's good) is 4. It follows that 4 is ingredient (2) of each minimum. Ingredient (1) is, per favoured option, others' good (that is, Rose's good) in that option: it is 2 if option 1 is favoured, 4 if option 2 is favoured, and so on. We can now identify the minima in the four sets and mark them with a circle. So for each option, the number in the circle is the situational virtue value of favouring that option.

We are now in a position to check whether these situational virtue values live up to the hopes that we placed in the hybrid theory. Suppose

situational virtue value of favouring a world
$= \min\{$others' good in the favoured world,
others' good in the world with the
largest sum others' good + own good$\}$

world no.	others' (=Rose's) good own (=Mary's) good	others' good + own good	
1	R—2 0⤬	2	$\min\{②, 4\}$
	efficient sacrifice		
2	R—4 −1	3	$\min\{④, 4\}$
	mere increase of own good		
3	R—4 −1	$\boxed{5}$	$\min\{④, 4\}$
	inefficient sacrifice		
4	R—5 −999 —···	−994	$\min\{5, ④\}$

Table 1: An example of a situation involving Mary and Rose shows how the theory of hybrid virtue delivers situational virtue values that behave as desired. The second column shows for each world first Rose's welfare (marked with an »R«), then Mary's; depending on which world Mary favours, her situational virtue value, defined as the minimum of two values, is the number in the circle.

that Mary favours option 2 instead of option 1. In other words, she is prepared to give up 1 unit of her own happiness for 2 units of Rose's happiness. That is an efficient sacrifice, a move with a good rate of exchange between her own good and the good of others. Here the situational virtue value is the sum of others' good and increases with that sum (viz., from 2 to 4). We wanted that increase. It reflects that it is virtuous to efficiently care for others.

Suppose that Mary favours option 3 instead of option 2. This amounts to favouring a mere increase of her own happiness. Since the situational virtue value continues to be the sum of others' good in the favoured world and the sum of others' good in the favoured world does not change, the situational virtue value does not change either, but

stagnates (at 4). We wanted that stagnation. It reflects that self-love is, within certain limits, neither virtuous nor unvirtuous.

And now suppose that Mary favours option 4 instead of option 3. In other words, she is prepared to give up 1,000 units of her own happiness for just 1 unit of Rose's happiness. That is an inefficient sacrifice, a move with a bad rate of exchange between her own good and the good of others. Here the situational virtue value is no longer identical with, and rises no longer with, the others' good. The situational virtue value stagnates (at 4), because the capping takes effect. We wanted that stagnation, too. It reflects that caring for others at an unproportionally high cost to oneself is neither virtuous nor unvirtuous.

Compare these pronouncements to those of the competing theories. The theory of virtue as self-including benevolence would claim, implausibly, that virtue increases when option 3 is favoured instead of option 2 – that is, when Mary favours more good for herself. The theory of virtue as self-excluding benevolence would claim, implausibly, that virtue increases when option 4 is favoured instead of option 3 – that is, when Mary favours an inefficient sacrifice. The theory of hybrid virtue that we have arrived at avoids such implausibilities and succeeds in capturing virtue better than the two blunter theories do. We are relieved from the pressure that was diagnosed in section 2: the pressure to commend an agent either for taking into account her own good to some extent; or, vice versa, for taking it into account less than that of others or even not at all. There is a clear and satisfactory way to praise benevolence without praising any of those features.

6. Conclusion

Three things have happened in this paper. I identified as clearly as I could a problem about the weight of self-love in benevolence and in conceptions of virtue as benevolence. Roughly speaking, the problem with regard to virtue is that the two enticingly simple accounts of virtue as benevolence – virtue as self-including or as self-excluding benevolence – are implausible in that they make morality either too welcoming or too hostile to prudence.

Next, I prepared at some length the solution of this problem, and of others that lie outside the scope of this paper, by introducing a framework for theories of virtue as benevolence. The framework makes use of the supervenience of virtue on situational preferences. It helps us

understand the structure of virtue and benevolence, and it permits us to express, study, and compare competing theories of that structure. In the presentation of that framework, the unsatisfactory claim that virtue is self-excluding benevolence served as an example. I finally proposed, within the new framework, a theory of hybrid virtue. That theory avoids the infelicitous claims of the overly simple conceptions of virtue as benevolence. It denies both that you increase and that you decrease your moral worth by taking into account your own welfare.

The theory of hybrid virtue may amount to good news for the virtuous person we encountered in the introduction to this paper, the virtuous person wondering whether to enjoy life a little more. The news is that there is leeway. Some variations in the strength of one's self-love do not affect one's virtue at all.