"This one," she said, pointing at a chocolate in the box she was handing to me, "is absolutely evil." And she was right or, at least, half-right: I've never tasted chocolate like that before, or since. Should I refrain from doing so?

Well, no, of course not. That chocolate—that is, the act of eating it—wasn't really evil, after all. It was very pleasurable and, we can reasonably suppose, innocently so. And even if it's not true that I have a moral duty to—should—engage in pleasurable acts, what is strictly true is that I don't have a moral duty not to do so. Provided that my doing so is indeed innocent, provided that it isn't undertaken at the cost of foregoing acts that I do have duties to perform, it seems perfectly okay. Pleasurable acts that respect that proviso are a subset of the set of acts that are neither obligatory nor forbidden. So why was she at least half-right in describing the act of eating that chocolate as "evil"?

Before we can get near to answering this question, a fair bit of ground-clearing needs to be done. Probably most philosophers would agree that—even leaving aside great chocolate experiences—we use the word 'evil' very unstintingly, perhaps too unstintingly. For one thing, it's used to describe not only acts, but also persons, dispositions and states of affairs. Moreover, on many conceptions of evil, some of these kinds of uses lack independence from the others: some are derivative from, or presuppose, one or another of the others. Thus evil acts, it's sometimes suggested, are necessarily the products of evil dispositions; or the evilness of states of affairs is a sufficient condition of the evilness of the acts that bring them about; or having evil dispositions is a necessary, and maybe sufficient, condition of being an evil person.

Again, on some conceptions of evil, evil is simply synonymous with wrong, bad, disvaluable, and other terms of negative moral appraisal. Other conceptions alternatively construe evil as what some have called a

“Calibrating Evil” by Hillel Steiner,
“wrong-intensifier.” And this construction is itself open to (at least) two significantly different interpretations. On the one hand, it’s taken to mean that the members of the set of things that are evil possess the very same properties as members of the set of things that are wrong, bad, disvaluable, etc., but they possess them to a degree which exceeds some (high) quantitative threshold. To intensify is often simply to increase: in this case, to increase the wrongness, badness, disvalue of that wrong thing. So, on this view, evil things are straightforwardly construed as very wrong ones. The wrongness of an assault increases with the extent of the injury intended and/or inflicted and, beyond a certain amount of injury, that assault becomes evil.

Alternatively, intensification can denote the addition of a qualitative difference. The wrong is intensified by the addition of some property that, in association with the wrong-making properties of the thing involved, renders it evil. This is a conception of intensification as aggravation. In law, an assault is an “aggravated assault” if it’s committed with intent to murder or with the aid of a dangerous weapon (and regardless of whether the injuries intended and/or inflicted are more extensive than those from assaults which lack these concomitant properties). The wrongness of your eating a sumptuous meal under the gaze of starving children is, we might think, aggravated by that food’s having been confiscated from them.

Finally, it’s a feature of virtually all conceptions of evil that they concur in attributing quantitative variability to evil itself. One evil act can be more evil than another. The familiar notion of “the lesser of two evils” is intelligible. Evil, in short, can (in principle) be calibrated.

The aforesaid ground-clearing consists, then, partly in this. It seems to me that there is some non-negligible demand, reflected in ordinary usage, for a conception of evil possessing the following four properties: (i) that it is not simply synonymous with other terms of negative moral appraisal; (ii) that it applies independently to acts, without logically committing its users to the existence of any connection between the evilness of those acts and the evilness of either their perpetrators or their perpetrators’ dispositions or the states of affairs resulting from their perpetration; (iii) that it is a wrong-intensifier in the aggravating or qualitative sense,1 and (iv) that it admits of quantitative variability. I can’t pretend that this conceptual demand, strong though it is, enjoys a monopoly license from
ordinary usage: precisely that it doesn’t has been the burden of the preceding paragraphs. Yet it seems indisputable that there is such a demand. And what I aim to do here is to explore the features of such a conception and, thence, their implications for assessing the extent to which evil acts are evil.

Various conversations that I’ve had with several people on this subject have indicated that a fruitful starting point for such an exploration is often thought to lie, oddly enough, in certain features of the supererogatory. ‘Oddly enough’ because, whereas evil acts are commonly taken to be especially bad or wrong ones, supererogatory acts are understood as especially good or right ones: the former contribute disvalue to the world, while the latter contribute value, often much value, to it. What two things could be more dissimilar?

It is, however, the very terms of this contrast that go some intuitive distance in sustaining the thought that we could usefully begin our search for a conception of evil with the aforesaid four properties, by looking at the supererogatory. For these terms suggest some symmetry between them that I’ll call the Negative Counterpart Thesis (NCT): that evil acts are simply the negative counterparts of supererogatory ones. There’s something about both of them—some shared set of features—that makes them distinctive-ly different from merely (sic) wrong and right acts, respectively. By closely scrutinising the possible grounds for NCT, we might hope better to illuminate the nature of our desiderated conception of evil. As we shall see, NCT is correct on one interpretation of it, but not on another.

Why, then, might we suppose that NCT is correct, that there is some symmetry between evil and supererogatory acts? In what sense do the latter possess the positive counterparts to properties (i) through (iv) What is the affinity thereby implied? The more obvious answer lies, I think, in the contrast described in the penultimate paragraph above. NCT appears to be invoking the idea of an ordinal scale somewhat analogous to the interval scales we use to measure temperature. There are points on those latter scales that are cold (very cold, etc.), points that are hot (very hot, etc.) and a range of points—or at least one point—between those two indefinitely large sets of points, that is neither hot nor cold. So, symmetrically, there are act-types which display varying degrees or levels of badness, ones which display varying degrees of goodness, and ones which do neither: that is, act-types which are morally neutral or optional or indifferent. This all
seems fairly straightforward, and serves to capture the idea that evil acts and supererogatory ones respectively involve intensified disvalue and value.

But obviously it won’t do the job we need done, since the wrong-intensifying difference that we’re looking for is a qualitative, and not a quantitative one. The feature that here symmetrizes evil and supererogatory acts is simply a pair of thresholds, lying at some distance on either side of the moral-neutrality point on the scale. On the present suggestion, it’s those valuable and disvaluable act-types that lie beyond these respective thresholds that are deemed supererogatory and evil, just as temperatures respectively lying beyond the counterpart thresholds are vapourising and freezing ones. NCT, on this basis, fails to satisfy condition (iii) outlined above: that is, it fails to sustain a conception of evil as a wrong-intensifier in the aggravating or qualitative sense.

Perhaps, then, NCT is simply mistaken: a cul de sac in our quest for that conception, rather than a high road to it. Maybe the proposed symmetry with the supererogatory is just misleading. And, indeed, there is at least one quite good reason for thinking that it is misleading. This reason has much to do with why Kant, Bentham and others have rejected the very idea of supererogatory acts. We might call this consideration the paradox of supererogation. Obviously, a concept that is, in itself, paradoxical is not going to be the most reliable instrument for illuminating the elusive properties of another mystery-shrouded concept. So what is this paradox?

The paradox is best grasped by reference to the structural affinities that obtain between the deontic and the axiological structures for the moral appraisal of actions. In standard deontic logic, all act-types are exhaustively assignable to one or another of three modalities: forbidden, permissible, obligatory. This, however, needs some clarification since, strictly speaking, obligatory act-types form a subset of permissible acts.4 That subset relation entails that “nothing is both obligatory and forbidden (impermissible).” Thus, since nothing impermissible is obligatory, if p is obligatory, then p is not impermissible, i.e., p is permissible.

In order to secure a genuinely tripartite classification of act-types, deontic logicians have therefore proposed a division of the “permissible” modality into strong and weak senses of permissibility. An act-type is weakly permissible if its performance is unforbidden but obligatory. And it is strongly permissible if its performance is both unforbidden and un-obligatory: that is, both its performance and its forbearance are permis-
sible. My wearing red socks is, standardly, a strongly permissible act-type. I'll henceforth reserve the permissible modality solely for act-types that are strongly permissible. On this construal, then, act-types that are permissible would fall into the zone of what I previously described as moral neutrality or optionality or indifference. They are act-types that are neither right nor wrong, neither good nor bad. The deontic police simply don't care whether or not I wear red socks.

Axiological appraisal similarly deploys a tri-modal classification. Some act-types are valuable, some are disvaluable, and some—like wearing red socks—are neither. Shouldn't we therefore expect that the act-type membership of each of these three modalities would roughly correspond to that of their deontic counterparts? Isn't any action axiologically appraised as the optimal—best—one in a given situation going also to be deemed as the deontically obligatory one there?

Significantly, the answer is "no." Supererogatory acts—for instance, the actions of saints and heroes—are very clearly located in the set of act-types deemed valuable in axiological appraisal. They are valuable, usually very valuable, acts and, more often than not, constitute the best ones among the several performable options in situations of practical choice. Yet it is definitive of supererogatory acts that they are "beyond duty": that is, they are not members of the set of act-types that are deontically obligatory. Nor, of course, are they deontically forbidden. Should we therefore infer that, like acts of wearing red socks, they are morally permissible in the sense of morally indifferent? Do we really think that, morally speaking, it doesn't much matter whether, in some dire circumstance, I choose to act heroically or, instead, choose to wear red socks?

I'm going to take it that the judgement that this inference is utterly implausible, is one that needs no defence. Yet it leaves us with a paradox. How can actions that are so valuable never be obligatory? I have no answer to this question and am thus inclined to share the doubt, advanced by Kant, Bentham and others, of the coherence of the idea of the supererogatory—if that doubt is one denying that there can be any modal distinction drawn between dutiful and supererogatory act-types. Faced with an array of practical options, the one which morality requires to be chosen—which is obligatory—in that situation, is that act which is better than all the rest.

Nevertheless, and even if one does accept the existence of a disjunctively fourth modality of deontic moral appraisal—supererogatory, as well
as forbidden, permissible and obligatory act-types—we can readily identify the deficiencies of one interpretation of NCT. For whatever may be the contours of a conception of evil that satisfies the previously listed four conditions, it seems undeniable that evil acts are wrong or disvaluable ones: they are fully paid-up members of the set of act-types which are forbidden. As such, they simply cannot be construed as the symmetrical negative counterparts of supererogatory acts, if such acts are indeed unobligatory.

Let's call this mistaken version of NCT the modal interpretation. The symmetry it mistakenly purports to find, between evil and supererogatory act-types, consists exclusively in their locational properties within the modal structures of moral appraisal. That symmetry, it's true, does hold in respect of axiological structures: evil acts are disvaluable and supererogatory ones are valuable. But this axiological symmetry is insufficient to satisfy our aforementioned conditions (i) and (iii). For it supplies no reason to suppose that evil is not simply synonymous with wrong, bad, etc., and it similarly supplies no reason to suppose that evil is a wrong-intensifier in a qualitative, as opposed to merely quantitative, sense. Moreover, the modal interpretation of NCT sustains no such symmetry in respect of deontic appraisal. If, then, modal location cannot be the basis of NCT symmetry, what can?

Consider two cases previously mentioned above: the act of eating a sumptuous meal under the gaze of starving children from whom that food has been confiscated, and the acts of saints and heroes. In the former case, there seem to be at least two, if not three, properties of the act involved that make it wrong. It may be wrong to eat sumptuous meals at all, in a world with starving children. But it's certainly wrong to do so under their gaze and, again, certainly wrong to eat food that belongs to them. All of these different properties, we can readily agree, render that act wrong, indeed very wrong. But do they render it evil? The problem with saying that they do seems to be that this looks like wrong-intensification of a purely quantitative kind. Each of these properties, taken on its own, is already a wrong-making property of acts: the act-types they each signify are easily located in both the disvaluable and forbidden modalities of axiological and deontic appraisal, respectively. Ignoring the vital needs of others, tormenting them, and taking advantage of their unjust deprivation, are each pretty standard forms of wrongdoing. That one act incorporates all three of these forms of wrongdoing is surely sufficient to deem it very
wrong. But there seems no reason to suppose that such aggregation intensifies that wrongness in a manner different from that in which an act of bombing ten innocent persons possesses more wrongness than it would do if that bomb had hit only five innocent persons. Wrong-intensification, due to the addition of different wrong-making properties, doesn’t give us the conception of evil that we seek: it doesn’t give us condition (iii).

Something quite similar can be said about the acts of saints and heroes. In Charles Dickens’s *A Tale of Two Cities*, Sydney Carton takes Charles Darnay’s place on the guillotine. In acting to rescue Darnay from death, he acts to prevent the perpetration of a grave injustice, to ensure the future happiness of the woman (Lucie Manette) he loves, and to atone for the dissolute life he has led. Each of these intentions embodies what would, pretty uncontroversially, be deemed a right-making property of acts. Carton’s act is, so to speak, *very right*. But it might also be equally right—even had there been no Lucie Manette and even had he led an exemplary life—if his taking Darnay’s place had been an act to prevent the perpetration of *three* grave injustices. On this account, then, his (actual) act’s intensified rightness, due to its aggregation of several different right-making properties, does not give us a supererogatory counterpart to the qualitatively distinct conception of evil that we’re after. Yet, at least for those who accept the category of the supererogatory, Carton’s act is typically seen as a clear instance of it.

What is it, then, that could make the sumptuous eating an evil act and Carton’s act a supererogatory one? It seems to me that there’s no great mystery here: the notions of *self-indulgence* and, correspondingly, *self-sacrifice*, spring readily to mind. The sumptuous eater, we may suppose, is deriving some pleasure from his wrong act, while Carton performs his very right act with complete awareness of its severe and painful cost to his personal well-being.

If this analysis is correct, the basis of NCT symmetry appears to lie in the *affective* properties of evil and supererogatory acts. It’s the affective interpretation of that symmetry, rather than merely the modal version, that gives us a conception of evil that satisfies our conditions (i) through (iii). Evil acts are wrong acts that are pleasurable for their doers, while supererogatory acts are right acts that are painful to perform.10

What about condition (iv)—the requirement that our conception of evil be one that admits of quantitative variability, one that can be cali-
brated? I here simply assume that most accounts of morality embody some scale for distinguishing lesser from greater instances of right and wrong actions. Just what the properties of such a metric could be—whether they involve ranking or weighting or other modes of comparison—is a matter of considerable controversy and, fortunately, not an issue that needs to be addressed here.11 For whatever they might be, what seems to be straightforwardly implied by the foregoing argument is that any such scale needs to be combined with something like a felicific calculus, in order to assess how evil an evil act is. That is, the scale of evilness is a compound scale. Like the metric for measuring the density of a substance (weight/volume), the scale for the measurement of evil would thus be one combining the scale for wrongness with that for pleasure.12 An obvious implication of such a scale is that there can be acts which are very wrong but not very—if at all—evil. And conversely, some highly evil acts, though they must be wrong ones, may be only marginally so.

Some persons may find this counter-intuitive.13 One way, perhaps the only way, of assessing counter-intuitiveness here, is to test that implication against real or hypothetical cases. Thus we might consider two Nazi doctors performing similarly gruesome experiments on similar numbers of Auschwitz inmates. One of them gains little pleasure from the acts he is performing, while the second thoroughly enjoys them. Now, however we calibrate both the considerable wrongness, and the evilness, of what they are doing, it seems reasonably clear that our assessment of the second is bound to be the more adverse of the two—or so I shall assume. On my proposal, then, that difference consists entirely in a difference between the degrees of evil we associate with their respective sets of actions. The wrongness of what they each do—their respective contributions of disvalue or the forbiddenness of their respective activities—seem to be the same: it’s not the disvalue or forbiddenness of the second’s activities that we would regard as being reduced if, every time he entered his laboratory, he were to develop a headache that diminished his enjoyment of what he was doing. So if our assessment of his conduct is indeed more adverse than that of his colleague’s, that difference looks to be distinctly attributable to a difference of evilness, in the sense of evilness developed above. It’s because he gets more pleasure than his colleague, from doing the same wrong thing, that his actions are the more evil ones.14

By way of a conclusion, let me just say this. I do appreciate that those who have wrestled long and hard with the perplexing idea of evil may
CALIBRATING EVIL

well find the foregoing analysis of it—particularly in regard to its calibration—somewhat shallow and mechanical. Evil is indeed a rich and complex concept and, as was noted near the outset, the conception of it presented here certainly holds no monopoly license from ordinary usage. Yet the challenge posed for those who do find this account unacceptable is to identify an alternative conception of evil’s properties that nonetheless manages to satisfy conditions (i) through (iv) since these are, I think, reasonably well-entrenched in at least part of that usage.

So the reason why she was only half-right in describing the act of eating that chocolate as “evil” is simply that, although it was undeniably pleasurable, it wasn’t at all wrong.15

Hillel Steiner

University of Manchester

NOTES

1. And, as a wrong-intensifier, it can in principle pertain to any kind of wrong act and not solely interpersonal ones: wrongs to other persons are a subset of the set of wrong acts.

2. It’s worth noting, at this point, that the history of moral philosophy reveals several highly influential meta-ethical positions that reject the very idea of supererogatory acts. Kantian deontology and Benthamite consequentialism simply assimilate putatively supererogatory acts into the set of acts that are right or obligatory. The grounds for this rejection and assimilation will emerge presently.

3. ‘Indifferent’ in some absolute sense, as well as in the (economist’s and others’) comparative sense of ‘mutually indifferent’ or ‘equivalent’. A pair of good (or bad) act-types can be morally indifferent in this latter sense. Act-types which are absolutely indifferent are also mutually indifferent.


7. I here put aside one complication that has no bearing on the present issue. The available options in a situation may all be instances of act-types that are deontically forbidden: When asked for my opinion of his work by an anxiously aspiring artist, I may have to act either dishonestly or unkindly. This possibility—of having to choose between two wrongs—arises even in situations where some of the available options have right-making properties, since that fact in no way precludes all of them from also (as in the present case) having wrong-making properties. This is not, however, usually taken to
signify that none of these options can be deontically obligatory and that the choice to be made by the person confronting them is, like a choice of socks, a morally unconstrained one. A rational moral code—one whose injunctions cannot generate contradictory judgements in particular cases—is one that satisfies, inter alia, the condition of completeness: it supplies a strong ordering of all available options; cf. my An Essay on Rights, (Oxford: Blackwell, 1994), pp. 111–25. And non-contradiction looks to be a plausible requirement for any set of propositions, including those implied by normative systems. Thus Risto Hilpinen, Deontic Logic: Introductory and Systematic Readings (Dordrecht: D. Reidel, 1971), p. 16, says: “The principles of deontic logic determine conditions of consistency for normative systems. By a ‘normative system’ we understand here simply any set of deontic sentences closed under deduction. When is a set of deontic sentences consistent? It seems natural to require that at least the following ‘minimal condition’ should be satisfied: . . . [that] all obligations in this set can be simultaneously fulfilled, and that [an act] is permitted only if it can be realized without violating any of one’s obligations.” This is equivalently true under axiological appraisal, for situations where all the available options are instances of disvaluable act-types. Creating a mistaken impression of my views in someone’s mind, and causing him to suffer a loss of self-esteem are, we may suppose, both disvaluable act-types which axiological codes must order, to satisfy the completeness condition.


9. What if all that morality requires is, rather, that we satysfice, that the action we choose be one that is (merely) good enough? After all, isn’t it true that “the best is often the enemy of the good”? Adequately addressing this complex issue here would I fear, take me too far afield from the central focus of this essay. That said, it seems to me that the plausibility of such a view strongly depends upon the reasons offered for choosing only a good-enough action, not being ones that themselves invoke moral considerations. These reasons standardly cite cost as the factor justifying sub-optimal choice. And that cost is usually associated with either (i) a delay in choosing due to the difficulty of identifying the optimal action, resulting in an even more sub-optimal outcome, or (ii) a serious compromise of the choosers interests. Of these, the first clearly is a moral consideration, signifying no more than that time- constraints, like many other types of constraint, strongly influence what counts as the morally optimal action: ought implies can. And the second suggests only that prudence is itself morally valuable: that regard for the choosers interests is one amongst the several factors entering into the identification of the optimal action. In neither case are we driven to conclude that morality’s requirements are satisfied by a sub-optimal action.

10. Pursuing this line of thought, a plausible suggestion is that the sets of evil and supererogatory acts can be enlarged by weakening the affective requirement for acts that are members of them. Thus an act is evil if (a) it is wrong, and (b) its doer does it either pleasurably or with affective indifference; correspondingly, an act is supererogatory if (a) it is right, and (b) its doer does it either painfully or with affective indifference.

11. The requirement of completeness in a normative system (see n. 7, above), though presupposing the presence of some such scale, is indeterminate with respect to the metrical properties of its modes of comparison.

12. And, appropriately conversely, rightness/pain for the measure of supererogatoriness.
13. Apparent counter-examples are the acts of joy-riders and computer hackers and, conversely, those of saints who enjoy the things they do for others. Isn’t it thus mistaken to regard the former as being evil and the latter as not being supererogatory? My guess is that our disinclination to see joy-riders and hackers as evil is based on an impressionistic statistical judgement that the wrong they do is minor; that is, where we to have reason to revise that judgement upwards, the disinclination would, I conjecture, diminish. The saints objection is more troubling, though it does seem to presuppose what there is no reason to believe: namely that all the good acts of saints are supererogatory ones. In this regard, it’s not irrelevant to note both that martyrdom is not a necessary condition of sanctity and that it is, ceteris paribus, a sufficient condition of it.

14. An obviously important question is whether this conception of evil is capable of sustaining the widely held view that the Holocaust is paradigmatically evil. I think the answer can be “yes.” The Holocaust was not a single act, but rather an agglomeration of countless acts performed by a sizeable number of actors. As the foregoing example suggests, it’s unreasonable and, indeed, unnecessary to suppose that every one of those acts possessed the same affective properties, and to the same degree. All that’s needed to sustain that aforementioned widely held view, on this conception, is that a preponderant proportion of those agglomerated acts possessed the relevant affective—i.e., pleasure-giving—property to some degree. To describe a football team’s performance as magnificent (or abysmal) is not to impute that quality to each one of that performance’s constituent acts. See further my “Persons of Lesser Value: Moral Argument and the ‘Final Solution,’” Journal of Applied Philosophy, 12 (1995), 129–41, esp. pp. 135–36.

15. This essay owes a great deal to many conversations with Eve Garrard and Stephen de Wijze, and to extensive comments supplied by Jerry Cohen, Cecile Fabre, Maria Paola Ferretti, Adam Morton, and Peter Vallentyne.