

Article

A Puzzle about Self-Sacrificing Altruism

Saul Smilansky

Department of Philosophy, University of Haifa, Haifa 3498838, Israel;
smilsaul@research.haifa.ac.il

Submitted: 25 September 2020, accepted: 7 March 2021, published: 25 April 2021

Abstract: I present a puzzle concerning individual self-sacrificing altruism (SSA) that, to the best of my knowledge, has not been considered before. I develop an argument that challenges the common sense attitudes towards self-sacrificial altruism in typical, paradigmatic cases. I consider SSA involving sacrificing one's life for other human beings, focusing, for the sake of simplicity, on saving a single person. We have reasons to think that many paradigmatic acts of SSA may, on reflection, be irrational, that typical moral heroes are mistaken, that dispositional self-sacrificers should perhaps resist their good urges to keep saving people, and that the enchantments of heroism should regularly be resisted.

Keywords: self-sacrifice; supererogation; altruism; moral calculation; romanticism; moral paradoxes; crazy ethics

How to cite: Smilansky, S. A Puzzle about Self-Sacrificing Altruism. *Journal of Controversial Ideas* 2021, 1(1), 7; doi:[10.35995/jci01010007](https://doi.org/10.35995/jci01010007).

© 2021 Copyright by the author. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY 4.0) license.



1. Preliminaries

I present a puzzle concerning individual self-sacrificing altruism (SSA) that, to the best of my knowledge, has not been considered before.¹ I develop an argument that challenges the common sense attitudes towards self-sacrificial altruism in typical, paradigmatic cases. We have reasons to think that many paradigmatic acts of SSA may, on reflection, be irrational, that typical moral heroes are mistaken, that dispositional self-sacrificers

¹ The topic of supererogatory self-sacrifice has been relatively neglected: "Despite this important role that sacrifice plays in our moral thinking, moral philosophers have had surprisingly little to say about the nature of sacrifice" (p. 301). Marcel van Ackeren and Alfred Archer, "Self-Sacrifice and Moral Philosophy," *International Journal of Philosophical Studies* 26 (2018): 301–7. The main exception has been in discussions of supererogation, since J.O. Urmson, "Saints and Heroes", in *Moral Concepts*, ed. Joel Feinberg (Oxford: Oxford University Press, 1958/1969). See also, for example, David Heyd, *Supererogation* (Cambridge: Cambridge University Press, 1982); Gregory Mellema, *Beyond the Call of Duty* (New York: State University of New York Press, 1991); Alfred Archer, "Supererogation, Sacrifice and the Limits of Duty," *Southern Journal of Philosophy* 54 (2016): 333–54. Yet the puzzle I am raising concerning paradigmatic cases of individual SSA does not seem to have been noted.

should perhaps resist their good urges to keep saving people, and that the enchantments of heroism should regularly be resisted.²

I consider acts of SSA involving sacrificing one's life for other human beings, focusing, for the sake of simplicity, on saving a single person. I believe that the argument can be extended to saving more than one, but will not do this here. I assume that the self-sacrificing agents in question have good lives which are worth preserving, that they are happy about their lives, and that they could (if it were not for their self-sacrifice) expect to enjoy their lives for a lengthy period of time. The self-sacrifice is, for them, something that they would in itself wish not to do; it is a great *sacrifice*, and they do it at least in part because they hold that it is the right (good, correct, noble) thing to do. Some instances of SSA are decided upon on an impulse, others allow for more time for reflection but, even in the former cases there is an awareness of the nature of the act. I also assume that there are no motivational pathologies involved in the sacrifice.

I focus on the objective goodness or rightness of acts of SSA, and assume for the sake of this discussion that the idea of objective goodness or rightness makes sense (a number of different meta-ethical views might suffice here). I look at SSA that is voluntary, intentional, autonomous and supererogatory (that is, beyond one's duty). The people saved are strangers to the self-sacrificing rescuer and the latter is not acting in any relevant professional capacity, nor is the saviour otherwise responsible for saving them. I also assume that the self-sacrificer's life is "for her to give", in that she is not, say, a single parent to five young children, so that self-sacrifice would involve great direct harm to others to whom she has major obligations. I concentrate my efforts on instances of SSA that involve sacrificing one's life, or putting it at great risk. The self-sacrificers need to be conscious of the nature of their decisions and actions as instances of altruistic self-sacrifice of this sort, at least in a weak sense. Yet putting one's life at what one reasonably takes to be great risk suffices for the sake of our discussion, even if the hero recognises the possibility of survival and indeed survives. In the rest of this discussion, we shall speak of instances of SSA as genuine instances of self-sacrifice, even when there were great risks taken that did not in fact materialize; and hence it was not an actual self-sacrifice in the strict sense. I also assume, for the sake of this discussion, that those whom the would-be saviours attempted to save were indeed saved, and that it was reasonable for the would-be saviours to assume that they would be. We do not need to concern ourselves here with doubts about committing an SSA that follow from the chance that the person who is to be saved will not in fact be saved.

Finally, I focus on individuals sacrificing themselves as individuals, not as part of collective groups or within special structures (for example, army conscripts). These cases raise questions concerning matters such as legal obligations, contracts, and collective goods that would take us in different directions.³

² If convincing, this would merge with a broader view concerning morality which I call "Crazy Ethics" (first introduced in Saul Smilansky, "Free Will as a Case of 'Crazy Ethics'," in *Exploring the Illusion of Free Will and Moral Responsibility*, ed. Gregg D. Caruso (Lanham, MD: Lexington Books, 2013), and with my emphasis on the largely neglected importance of paradoxes concerning ethics and the meaning of life in Saul Smilansky, *10 Moral Paradoxes* (Malden, MA: Blackwell Publishing, 2007); Saul Smilansky, "Paradoxes and the Meaning of Life," in *Oxford Handbook for Meaning in Life*, ed. Iddo Landau (Oxford: Oxford University Press, 2021, forthcoming). For a different challenge to common-sense concerning sacrifice, see Saul Smilansky, "Should We Sacrifice the Utilitarians First?" *Philosophical Quarterly* 70 (2020): 850–67.

³ See, for example, Joseph Raz, "On the Moral Significance of Sacrifice," *International Journal of Philosophical Studies* 26 (2018): 308–14.

Paradigmatic examples of SSAs would include jumping in and trying to save another person from drowning in a turbulent sea, an action that is extremely risky given one's swimming ability; confronting, bare-handed, criminals who are tormenting an innocent person in a subway; or jumping in front of a running horse that would otherwise trample over a child. Such acts are typically seen as being good yet much beyond reasonable moral expectation, and are thus supererogatory and highly admirable, indeed heroic.

The basis for our discussion will be, as stated, the idea of supererogation. This means that we will be setting aside strong consequentialist positions (such as traditional utilitarianism) which rule out the supererogatory. The idea of the supererogatory also means that, conceptually, the cases involved will not be ones in which deontological duties to risk one's life in order to save others will apply. Likewise, we will be assuming the absence of any tacit or explicit contractual arrangements here; nor are there any pertinent expectations of reciprocity. This means that the traditional normative theories will not be the direct way to evaluate this supererogatory altruism puzzle. Hence, setting aside empirical uncertainty, the evaluation of instances of SSA will aim for broad sensitivity to the emerging gains and losses in value. Whether an instance of SSA is correct or mistaken shall primarily depend on whether the act increases or decreases value. The aim will be to try to capture at least initially a broad spectrum of intuitions.⁴

Let us now consider a few examples that can help us understand the distinction between mistaken and correct acts of SSA. The easiest sort of case would be the dramatic risk-taking involved in rushing into one's burning house when it is going up in flames, trying to save the family's pet (say, the pet hamster), who is caught inside. For all the sorrow about the hamster, one should not put one's life at great risk for such a cause. Let us call such examples *Hamster cases*.

Assume now that two young men and a young woman sacrifice themselves in order to save one other person who is known to be a no good, hate-filled, despicable, and previously violently harmful bully. All decent people who know this person think of him as a worthless scum. Let us assume that he is no longer actually dangerous, despite his continuous vile racist and sexist desires and intentions. Let us call him the *Despicable Bully*. It seems to me that there are excellent reasons to think that these two men and the woman have made a *mistake*. Together they have sacrificed themselves in order to save a person who is much less worthy. Even if we were not to go so far as to say that the world is better off without him, he is manifestly not worth the sacrifice of the lives of three good people. This result seems to be sustained also if there is only one person who sacrifices him or herself.

A typical case of SSA, by contrast with that of the bully, is thought to be very different. In a typical case, one sacrifices oneself and saves, we assume, a fairly good, decent human being. If one risks one's life and saves such a person, this is considered a great deed.

⁴ The rejection of impersonal comparisons of states of affairs in terms of value would be a way of avoiding my conclusions, but this is an intuitive price that very few will be willing to pay; see *David Heyd, Genethics: Moral Issues in the Creation of People* (Berkeley: University of California Press, 1992); Richard Kraut, *Against Absolute Goodness* (Oxford: Oxford University Press, 2011).

2. The Case against SSA

2.1. “Counting Oneself as a Hero”: Including One’s Own Exceptional Worth, as a Hero, in the Moral Calculation

Another person was about to die, and instead will live, because you, in an act of SSA, sacrificed yourself in order to save him. So, the result is a “switch”: instead of a situation in which that person dies and you continue living, we have one in which he lives but you die.

What has been left out of the evaluation, however, is the value of the self-sacrificing altruist, *once she has done her deed*. For, in the process of sacrificing oneself for the sake of others the world loses more than just an ordinary person, of ordinary importance or worth. It loses a morally rare, unusually superior, human being—*one who has altruistically sacrificed herself for the sake of others*. And this begins to look like a much more dubious proposition, and quite possibly a mistake.

Recall the good people who sacrificed themselves for the sake of the *Despicable Bully*; it was compelling to think that they made a mistake. Similarly, sacrificing oneself for a person of ordinary moral worth, who (we can assume) would *not* sacrifice himself for others, in such a situation, or manifest *any* similar moral quality throughout his life, is arguably a similar type of mistake.

The stark choice appears to be between (i) being a hero and making a moral mistake; or (ii) not attempting to save the other person. And being a hero is a mistake *because* it makes one a hero, who in that very process of becoming a hero is being sacrificed.

If this is compelling, we have here an intriguing problem:

1. Acts of SSA that save only one person are a wonderful thing (they are morally amazing, supererogatory acts of altruistic self-sacrifice).
2. However, *because* they are so wonderful, acts of SSA (at least when they save only one person) are arguably always mistaken.
3. Acts that are always morally mistaken are not morally wonderful.

This might be seen as a paradox, or an elegant *reductio* of the act of SSA. We need not insist on this strong result throughout the SSA range; the striking nature of the problem will be sustained even with weaker assumptions, and even if the evaluations do not apply to all cases.

Hence, *altruistically sacrificing oneself in order to save another single human being is a morally dubious act, and quite possibly a moral mistake*. By doing so, one thereby sacrifices a person who, through that sacrifice, has (in nearly all cases) become much more valuable than the person one saves.⁵

Recall that we are assuming throughout that it is supererogatory to save the other person, and one would not be bad were one not to do it. There is, therefore, no symmetrical paradox here, such as that one is bad for *not* taking up the opportunity of becoming an SSA.

⁵ We note but need not concern ourselves here with the distinction (familiar, for example, from discussions of utilitarianism) between the standards that make an action right, and the justified decision procedure. It is one thing to say that a feature may render an action wrong, it is quite another to say that we always should consider that feature. Our focus is on the conclusion that emerges from adequate calculations, calculations which determine the moral status of the self-sacrificing action.

Let us return to the decision to opt for SSA. Before the act of SSA is done, it might seem as though it is a good idea to do the “switch”, saving the other at one’s expense, or at least not a bad idea, because you and that person are, still, of more or less equal value. However, if one understands the point of this paper, then, when one is about to make the sacrifice, one already understands that, once the deed is done, it quite likely will have been a mistake, for one has, overall, destroyed more value than one has saved. SSA can appear to be a moral opportunity but, on reflection, it suddenly begins to seem, very possibly, like a moral mistake.⁶

It might seem puzzling where the lost value of the self-sacrificer is. Once she has done her deed, she ceases to exist. So there is no person remaining to whom the value would be assigned. But this is a false puzzle. The very problem is that we have lost an unusually morally admirable human being. The value of the person who is no longer in existence has been lost, and *that* is what is bad about the ensuing state of affairs.

The significance of counting oneself as a hero in this way is further supported by the idea of *redemption*. When a person has great matters to morally redeem, such as that he has intentionally harmed others in large and manifestly immoral ways, then self-sacrifice for a morally good cause seems much more acceptable. Consider *Mafia Guy*, a repentant former murderous Mafia operative, who tells all to the police, knowing that he is very likely to be killed for doing so. Since he has greatly *reduced* his moral value through his past immoral acts, he sacrifices less, when sacrificing oneself; he begins with less value, *is* a person of less value, than an ordinary person. Being quite properly ashamed of his past actions reflects such a lesser evaluation of himself, as compared to the norm, and a wish to redeem himself if he only can. By taking the great risk of self-sacrifice, then, he does not destroy the greater (a self-sacrificer) for a lesser (an average person), but greatly adds value, through the self-sacrifice, to a deeply defective human being. Under certain circumstances, through the act of SSA, the self-sacrificer may gain sufficient value to reach, more or less, a moral equilibrium. One sacrifices a person with a large “moral overdraft”, which is then overcome or redeemed by the sacrifice.

There seems to be a direct relationship here between desert and moral worth. This person’s freely sacrificing himself in the way we are considering makes him deserving; he acquires unusually high moral worth and has become *praiseworthy*. This moral worth can in some measure redeem his past misdeeds, compensating for his past blameworthiness. Note that we needn’t understand what’s going on here in terms of a single dimension of value that has gone down and can be raised up. We might see here a loss of a certain default “baseline” status, that can perhaps be partly regained via sacrifice. I am using a consequentialist framework of comparative value, but much of my argument can be sustained even without it.

Note also that the numbers seem to strongly affect how we think about SSA cases—risking death to save 30 ordinary people seems more easily justified than if only one such person is at stake, which suggests in another way that even common-sense is

⁶ Another, consequentialist, argument against SSA of the sort under consideration might be that people who are prepared to sacrifice their lives in order to save another are likely to do more good in the world (apart from the sacrifice) than the average ordinary person, in the future. Hence their remaining alive, rather than the (we are assuming) ordinary person they might save, would be preferable. This argument however seems vulnerable to doubts such as whether a potential SSA would indeed do significantly more future good, and to opposing consequentialist considerations, such as the collateral effects (see below) of the example set by the self-sacrifice.

concerned with the value of the saved. For this reason I have in this paper framed the puzzle as SSA applying to saving a single person.

My claim that typical cases of SSA involve sacrificing a person who, through the sacrifice, has become much more valuable than the person she saves, raises a number of questions. The first is whether the self-sacrificing person indeed gains in moral worth only when she sacrifices herself. If she is strongly disposed to sacrifice herself for moral reasons should the occasion arise, is she not already of unusual moral worth—more valuable in the relevant sense than others who entirely lack such a disposition? There are various possibilities here. One is that it is only with the act that the person gains in value. Another is that she already has some high value just because of her character and dispositions, but gains even more in value when she actually does the self-sacrificing act. A third is that actually doing the act does not matter. I will assume in the rest of the discussion that all or most of the value arises with the act of SSA. It seems to me, for example, that, although one is already admirable for being willing to consider SSA, if one then stops short of doing so, one's value remains much lower than it would have been had one acted.⁷ But my position does not require that we decide here. For, after all, the question whether acts of SSA are mistaken is concerned with *the acts*.

A second issue is what kind of value this is. It is not primarily the value of her life *to her*—that is, it is not that it makes her life more worth living (although, if she survives, some such factor might come into play, for sacrifice makes one's life more meaningful, and that meaning is not just impersonal). It may be partly, but not exclusively, her value to others. But it seems to me to be primarily *impersonal value*. Persons who engage in morally admirable action gain in impersonal value. The world is better for their existence. If the act is admirably heroic then, even if the person survives and continues to live much longer, the value she has gained is retained (more on this below).

2.2. A Further Paradox: *The Past-Virtuous*

Being a person who is ready for SSA makes one morally rare and worthy. We have seen that it is compelling to take this into consideration when contemplating such acts of SSA. However, what does this imply for those persons who have *already* been ready to do acts of SSA, and attempted it in the past? For example, consider a woman who saved a drowning child, and did so while severely risking her own life, being ready to sacrifice her life in the deed. In such a case it seems difficult to deny that she is a moral heroine (even if she made a mistake by the self-sacrifice; see below). Such a distinction remains for life; hardly anyone can boast as much. It would still be correct to refer to her as a moral heroine even a dozen years later, and even if she does not perform any further acts of heroism (albeit perhaps only if she does not display in those years significant moral cowardice or acute actual egotism).

What does this mean? If we follow our previous conclusion, then this woman needs to be extra careful. *Being so valuable, she has to take great care not to put herself at risk*. Contemplating any further acts of SSA heroism of the usual sort (for example, risking her life to save yet another drowning person) should become quite unthinkable. Arguably, she should let the second person drown. Sacrificing herself (when she is already a hero of the SSA sort) would appear to be a *mistake*. Not only would she be sacrificing a moral

⁷ The question concerns altruism in general, and we can set aside here the doubts that follow from the present puzzle.

hero, but (in this second act) a repeat hero; something still more noble and rare. As a rule, following this path implies that there should not be repeat heroes of this sort, nor continuous ones. Heroes should stop being (active) heroes. *The more heroic you are, the less you should contemplate heroism*, and indeed it is best if you are not too concerned with the threats to others that might require acts of self-sacrifice of you. However, at least initially, this seems *absurd*, and sounds more like a *reductio* of the *Counting Oneself as a Hero* view than a compelling argument.

It might be countered, first, that the one-time hero surely is not morally amazing in every way, and indeed she can have many moral faults. Second, that there is no special, rare value in self-sacrificing altruists, and in the concomitant importance of their continuing to live in the world, unless they *continue* to act in SSA ways (when presented with opportunities).

But focusing on the faults of a SSA hero when considering her overall value; or dismissing her status as a person of rare value, unless she continues to exercise it at risk of self-sacrifice, is deeply mistaken—and unfair. Compare the idea of a genuine war hero: surely one retains the status even if the war was long ago, and one has chosen to live a mundane, average life ever since. SSA of the right sort, as we are assuming in our discussion (that is, non-pathological, intentional, immense risk-taking, and so forth), needs to be considered as giving the hero *great* and *continuing*, if not permanent value.

3. Objections

In response to the apparent paradoxes we have encountered, of *Counting Oneself as a Hero* and *The Past-Virtuous*, and the threatened crumbling of much of the ethos of heroic SSA, one might veer strongly in a subjective, romantic direction, disconnecting decisions (and post-mortem evaluations) about SSA from calculations of anticipated value. This direction has a fairly strong resonance in common attitudes towards paradigmatic acts of SSA such as we have seen, and self-sacrificing heroism more broadly. The hero of the SSA sort, we might say, is exactly the sort of person who does *not* do calculations; to do them, and in particular consider him or herself in the calculation, would mean threatening the whole idea of the hero.

However, this radical shift to romanticism comes at a heavy price, of disconnecting our attitudes towards SSA from the typical sort of moral “calculation” and justification of moral thinking. It would, as it were, concede that we are unable to rationally defend typical acts of SSA. It would also distance moral action and thought from those places where “calculation” seems correct in indicating that SSA would be a mistake (for example, the *Hamster* or *Despicable Bully* cases). Similarly, we would lose touch with the situations in which SSA manifestly can make good sense, such as in cases of the redemption in *Mafia guy*. Is there no other way in which common-sense concerning SSA can be defended, without veering towards a radical, and largely irrational romanticism?⁸

It might be argued, as against the case that I have made, that an act of self-sacrificing altruism, by its very nature as SSA, *creates* value, it *adds* value to the world. The *Counting Oneself as a Hero* does take account of the new and overwhelming value created by acts of SSA but, the proponent of the current view will say, it takes it *in the wrong direction*. What matters is not that the act of SSA enhances the self-sacrificer, thus allegedly making

⁸ Another cost is that this romanticism involves the hero taking a different view of herself from that which it is rational for the rest of us to take of her.

him so valuable as to be unfit for sacrifice—thereby paradoxically self-defeating the value of the act of SSA. Rather, the value established by SSA is a value created in the world. The self-sacrificing agent has given away his life, and that act *adds great value to the world*. A world thought to be populated by cold calculators and self-preservers is seen also to have heroes who, by thinking of others and not thinking of themselves, and overcoming their fears and self-interest, redeem humanity, adding to our world this special, rare, type of value. The creation of that sort of value is what makes SSAs not only legitimate, but so laudatory.

The problem with this reply, however, is that it seems to beg the question. It simply *assumes* that SSA in a case such as we are considering is, all considered, a morally acceptable and admirable act—and therefore that it adds great value to the world, *in a way that suffices to overcome the destruction of the value of the person sacrificing herself*. Yet that is the very matter that my challenge puts under contention. My claim, we recall, is exactly the opposite, namely, that acts of SSA such as we are concerned with emerge, on reflection, to be apparent moral mistakes; they are wrong, they *reduce* value. *The more value we wish to attribute to the person for her heroic acts, the worse this makes her error, in sacrificing herself*. My move here is like those of jujitsu fighters, who use their rival's weight against them.

Note that I do not require that we deny that there is value being added to the world by the self-sacrifice. For example, a death occurring in non-successful altruistic self-sacrifice (where the person is not saved) is impersonally better, *in one way*, than a senseless death in a road accident, because at least heroism has been displayed. More broadly, the appropriate sort of sacrifice has intrinsic value.⁹ I have pointed out, however, that our common neglect of the special value of the sacrificer is morally blind, and that, without begging the question, we cannot discount the weight of the great loss of the person who sacrifices herself.

We might note here that there is widespread and special mourning when someone like Gandhi, Sakharov or Martin Luther King dies. We recognize that there has been an unusually great loss—not perhaps to any individual but to the world. Even when such a person is very old and no longer engaging in acts of great altruism, we are aware of a special loss—that a person of the greatest worth no longer exists.

Moreover, speaking about the beauty of sacrifice in this way raises various dangers. One is that the “logic” of the romantic view leads to risk-taking and value-destruction beyond reason and morality. Repeatedly, in the mass media we see heroes praised for risking their lives for a pet or an object; for example, jumping into a frozen river to save the family dog, rushing into a fire to rescue the cat, or confronting the burglars in order to save the family's jewellery. Romantic gestures such as these are indeed chivalrous and heroic. But they are also typically irrational, careless, and irresponsible. Second, an added risk is that, paradoxically, the more imbalance there is in the results of the sacrifice, the greater the added value may seem. Some might even think that the sacrifice is all the more saintly if the beneficiary is clearly *unworthy* of the sacrifice by someone so much more morally worthy. Third, there is the risk of something resembling quasi-fascism here, for conceptually the romantic view sees the need for sacrifice as an opportunity, and the

⁹ See Jörg Lösckke, “The Value of Sacrifices,” *International Journal of Philosophical Studies* 26 (2018): 399–418.

safe, secure life which requires no self-sacrifices as barren land, lacking beauty.¹⁰ On examination, the “adding value” view raises the suspicion of involving fetishism of the romanticism of sacrifice.¹¹

It would not do for the presenter of this view to fall back on the subjective intentions or attitudes of the self-sacrificers, however romantic. For, after all, such intentions and attitudes are present also in *Hamster* and *Despicable Bully* cases, but despite the undoubted moral heroism involved, these are clearly moral mistakes, which it would be wrong to do, and that do not confer the sort of uncontentional justifying value needed here. Notice, again, that I need not deny that there is value attained by saving the *Despicable Bully*, and even the *Hamster*. My sceptical conclusion concerning paradigmatic acts of SSA such as we are considering is surprising and radical, but my argument is based upon our common sense intuitions in cases such as *Hamster* and *Despicable Bully*, that the self-sacrifice is a mistake, because of the *comparative* value of those sacrificing themselves and those being saved. My claim is that a person who sacrifices herself for an ordinary person is making a similar mistake, because in the process of the self-sacrifice she acquires a higher comparative value that makes the sacrifice *prima facie* a mistake.

It might be further argued that one’s error in self-sacrificing oneself erases one’s status as a moral hero or heroine, hence, paradoxically, once again making one’s self-sacrifice permissible. But, at least in the clearer cases, that would be unreasonable. One is heroic, but quite possibly mistaken; mistakes should not be made, but do not eliminate the heroism—they make it misguided. Surely we can leave room for the idea that in many paradigmatic cases of SSA the sacrificer *is* morally admirable, even if my argument that such sacrifice should be avoided is convincing. In any case, the proponent of the objection would not really wish to deny the morally heroic nature of one’s acts (just because one has made a mistake), for this would then undercut the very claim that, through one’s act, one is introducing great value into the world.

Another possible reply to my challenge might be that the evaluation of acts of SSA should not be made according to a balanced evaluation of loss and gain, of the resulting state of affairs, but in the light of the idea of the *gift*. The self-sacrificing hero *annuls*, as it were, the weight—as a consideration against the sacrifice—of (a) his own value as a person with interests; and, by implication (b) the additional exceptional value he attains or reveals by becoming a heroic SSA. He decides to sacrifice himself, to *give himself* up, *irrespective* of these considerations.

In reply, we need not deny that this may be the hero’s phenomenology. It is also what we seem to be admiring in self-sacrificing altruists, the person’s willingness to be responsive to the needs of an “other” with the ultimate gift, risking his or her life; the agent’s selflessness. Yet we can nevertheless persist in wondering whether the

¹⁰ For a closely related discussion of the inherent tension between moral improvement and the conditions for the attainment of moral worth see Saul Smilansky, “Morality and Moral Worth”, in *10 Moral Paradoxes*. For a different form of the attainment of high altruistic moral worth, see Saul Smilansky, “Moral Accountancy and Moral Worth,” *Metaphilosophy* 28 (1997): 123–34. By contrast, for a cynical interpretation of much of the attribution of moral worth, see Saul Smilansky, “Moral Demands, Moral Pragmatics, and Being Good,” *Utilitas* 22 (2010): 303–8.

¹¹ It might be argued that some version of this suspicion could also apply to my view about the greater value of the heroic person. It would merely be that, on my account, the value is added to the person as well as the world. Admittedly, it is important to my view that the SSA person not be primarily motivated by the aim of increasing her own value. Yet beyond that my view is the best antidote against these dangers. My emphasis on “calculation” and aversion to inappropriate self-sacrifice counter these temptations of the romantic tendency surrounding self-sacrifice.

self-sacrifice is correct. That it is convincingly thought of as a mistake in cases such as *Hamster* or *Despicable Bully* would strongly support scepticism here as well. Even if the self-sacrificing altruist discounts himself, we need to say that he cannot be the final determiner.

It might be claimed that my position, that people who become of greater impersonal value because of their moral merit have reasons to preserve themselves, raises a specter of dubious implications. If these reasons are stronger than the reasons to preserve the lives of ordinary people, what does this imply about, for example, health care distribution? Should governments now be reviewing people's moral ledgers to determine which are most worthy of receiving priority in getting vaccines?

But I can maintain my view concerning the higher moral worth that SSA can generate, without this necessarily implying anything about proper social policy. There are excellent moral and political reasons not to publically allocate vaccines according to perceived moral merit, just as voting rights ought to be allocated equally. This does not imply that the loss of any human life is, impersonally, of equal weight, just as not all persons vote equally sensibly. Likewise, if we recall repentant *Mafia Guy*, his assassination is manifestly more destructive of impersonal value than would be that of a ruthless *Unrepentant Mafioso*; yet it does not follow that the latter's assassin ought to receive a longer prison sentence.¹²

Moreover, common sense intuition does often favor the idea that, given a choice, it is better to save a morally worthy or deserving person rather than a scoundrel. Some favor the view that all potential victims should have an equal chance of being rescued, but many do not. My assumption about the differential value of morally worthy and morally unworthy people is not, here, in radical conflict with a large part of common beliefs about equality.

I may be thought to be too quick to dismiss the worry about how my view generalizes—surely it is not just a question of social policy. In particular, doesn't the same issue apply to supererogation more generally, and to many altruistic acts? If we are both equally morally worthy, and I help you, I become more morally worthy, and thus deserve more than you, and, in anticipation, should I then not help you? I think that we need to firmly distinguish here between "normal" assistance and the issue of self-sacrifice. There is no motivation for limiting altruistic assistance to equals, and no inherent difficulty with generosity; there is little reason to see this as, in itself, problematic. With SSA, by contrast, the contributor is on course to eliminate himself (whether this materializes in the end or not). The question is one of *survival*, of continuing to exist. Whether such elimination can be morally justified, in the paradigmatic sort of case under contention, is crucially dependent upon *who* survives, and whether this is good or bad.

It might be further claimed that the seemingly irrational heroic acts of SSA do make sense, from a more strategic perspective. We also have to take into account *collateral gains*. Acts of SSA will make many others feel good about humanity and, in turn, inspire many of them to contribute as well, resulting in further contributions (though typically not life-sacrificing ones). This is so even if we consider the cases which appear to be romantic excesses. People are inspired by hearing about the self-sacrifice for one's beloved pet; and burglars are deterred by the thought of a householder's life-risking possible resistance; even when those are, on the local level, far too risky and highly irrational.

¹² The role played here by general social and political considerations as constraints on public policy can be seen when such concerns are largely absent, in one-time individual decision making. Deciding to save a person because of considerations of outstanding moral desert can be permissible, as I showed in Saul Smilansky, "A Hostage Situation," *Journal of Philosophy* 116 (2019): 447–66.

Moreover, the disposition to be ready for self-sacrifice often goes together with other, more moderate positive tendencies, and hence is worth cherishing.

However, there is, first, a question of the extent to which such collateral gains actually occur, and what weight we ought to give to them as compared to the other factors. Some temporary public “feel good” effect, or even a minor increase in charitable donations, does not seem sufficiently important.

Then there is, second, a risk of begging the question, but here pragmatically: if the public perceives most acts of SSA as morally laudatory, then this may generate collateral gains. But if, as the challenge I pose suggests, many such acts are moral mistakes, and this is recognized, then the perception of actual acts of SSA as widespread moral mistakes is much less likely to produce collateral gains.

It might then be said, because of the importance of the collateral gains, that the nature of acts of SSA as mistaken-in-themselves should not be publically disclosed, and ought to remain as it were esoteric knowledge. Put differently, the positive value of these acts lies not so much in themselves, but in the effects of the way in which they are perceived. But this move in effect concedes my challenge, and indeed gives it so much weight as to require hiding it. For it says that in themselves, cases of SSA such as we are considering *are* mistakes, but this should not be publicized because we may thereby lose some unrelated, more numerous but much smaller effects, which may depend on the widespread false perception of those acts of SSA.

Moreover, there is something deeply repellent in the idea that minor, cumulative, non-heroic gains make the difference and tip the balance in favor of acts of SSA. These are the cases in which typically a human being (and an exceptionally admirable one at that) sacrifices his or her life. Perhaps such acts of SSA do have their indirect uses, but this can hardly be the major consideration. In cases of rushing to sacrifice oneself, what seems to matter is the person one tries to save, and one’s own life. If my argument is convincing, we should also worry about the effects of heroic examples of SSA in motivating others to commit *similar* acts. Any collateral effects on others that do not involve SSA seem minor, and not the right *sort* of consideration in favor of SSA.

Finally, it might be countered that all humans are of *equal value*, and that even performing an act of SSA does not make a person more valuable. And that, if accepted, would solve the difficulty. Sometimes it indeed matters only that one is a human person. For example, when we want to say that every (adult) person merits a right to have one vote in democratic elections in his or her society. However, in our context it seems highly implausible that *only* this kind of value (which each person is born with) matters; and no weight may be given to any special worth, merit and desert that one acquires through one’s intentions and actions, however admirable.

Furthermore, an insistence that only the formal sense of equal human worth, which everyone is born with, matters, would make it impossible to reach the conclusions we saw as very plausible at the beginning of this paper. For example, that sacrificing oneself for the *Despicable Bully* would be a moral mistake; or, by contrast, the moral acceptability, and indeed fittingness, of self-sacrifice in cases calling for moral redemption (such as in *Mafia Guy*). For, after all, all these people equally meet the criteria of personhood.

A more moderate egalitarian stance may, however, be more plausible. There might be an explanation for our intuitions concerning cases such as *Hamster* and the *Despicable Bully*, that is nevertheless compatible with the rationality of self-sacrifice in the paradigmatic cases that I am challenging. A crude view might be that it is a mistake to sacrifice oneself for anyone—human or animal—who is below some threshold of worth, but acceptable and admirable to sacrifice oneself for anyone above that threshold. The

Hamster does not have adequate moral standing to begin with, while the *Despicable Bully* has forfeited or significantly reduced his. Any person over a baseline of basic decency would always be worth saving.

Another and perhaps more plausible variant might be gradualist, with an imprecise threshold: it would imply that self-sacrifice is a greater mistake the farther below the threshold the potential beneficiary is, and more acceptable and admirable the farther above the threshold the potential beneficiary is. Yet while helpful in advancing our thinking, such attempts do not seem to me sufficient to diffuse my challenge. The loss of altruistic self-sacrificers, who are exceptional people with extraordinary value, does not seem to be adequately addressed even by these formulations, insofar as they nevertheless seek to retain our commonsense intuitions on paradigmatic instances of SSA. This and many of my previous points can be seen from a thought experiment resembling John Rawls' *Veil of Ignorance*, which we might term the *Test of Hope*.

4. The Test of Hope

4.1. Who Was Saved?

Assume a typical situation of SSA such as we are investigating, where one person jumps into the turbulent sea, risking his life in order to save a stranger. We know that the heroic savior has drowned and that the person he tried to save, who would have drowned otherwise, was indeed saved thanks to the heroic savior who drowned. We also know that the person saved was one among five very different people, one of whom was then out at sea, but we do not know which one. All we can do is *hope* that the sacrifice of the hero was worthwhile. Now consider a number of possibilities as to the identity of the person saved:

The *Evil Swimmer* (a person who sets out to drown those attempting to save him)

The *Unrepentant Mafioso*

The *Despicable Bully*

An ordinary person

A previous SSA Hero

Clearly the sacrifice of the hero was a mistake concerning the *Evil Swimmer*. There is something particularly awful about this case, because the person saved was out to lure such heroes as the one who actually saved him, and drown them. The *Unrepentant Mafioso* has not tried to harm the hero, yet it also clearly seems a mistake to save such a consistently evil person at the expense of one's life. Our *Despicable Bully* is not of such massive negative worth, yet I trust most readers will share the intuition that the hero's saving him has been a mistake, in terms of the resulting state of affairs. In all three cases a person of exceptional value has been lost, and the person he sacrificed his life for was below the moral baseline, and unworthy of the sacrifice.¹³ So far these intuitions seem fairly widely shared. Among those three, presumably we would hope the person saved was, at worst, the *Despicable Bully*. Surely, however, we would further hope that the

¹³ We might be more amenable to such a sacrifice if it were thought that it would lead the person saved to radically change his ways, and perhaps perform similarly valuable acts. But this of course is very rare, and cannot rationally be a consideration in favor of the self-sacrifice.

person saved by the heroic sacrifice was an ordinary, decent human being, *rather than any of these three people*.

Would we hope that the person saved was a past hero of a similar nature, rather than merely an ordinary person? Such a hope seems to me quite reasonable, and this is significant in defending the value of the morally heroic person over the ordinary one. Otherwise, why would we prefer and hope that the person saved was of the former sort, rather than of the latter?

Note that I do not require that we think that hope is an automatic indication of rightness, and on reflection this would often clearly not be so. If we know that a certain billionaire is buying quantities of lottery tickets for this week's lottery, it seems morally appropriate to hope that some other, much poorer person wins rather than him. The poorer person's winning would be a better state of affairs, all considered. Yet this does not entail that we would be permitted to fix the lottery so as to prevent this billionaire from winning. The *Test of Hope* is not always an appropriate measure of rightness or moral permissibility, but seems to be a good way of exploring our moral intuitions as to the comparative goodness of states of affairs.¹⁴ Nevertheless, moving from the results of the *Test of Hope* into action would sometimes be plausible.

4.2. Who Died and Who Survived?

Now consider a somewhat different version of *The Test of Hope* in the case of the swimmer. This time we know that *either* the heroic savior has drowned *or* that the person he tried to save has; in any case one drowned and the other did not. However, we do not know *who* of the pair has drowned and who has not. What do we hope will be discovered, once the veil is lifted, and the facts emerge?

It seems to me that here as well the conclusion concerning the first three persons would firmly emerge, and probably even more strongly. We very much hope that the *hero* is the one who survived, rather than—probably in descending order in terms of the strength of the hope—the *Evil Swimmer*, *Unrepentant Mafioso*, and *Despicable Bully*. What about the ordinary person? If my case has been persuasive, we will hope that the hero is the one who survived, even if this means that the ordinary person has not, however much we regret that they could not both continue to live. What about a past similar hero, such as a woman who a few years ago saved another swimmer, at similar risk to her own life? Here my intuition is that I am at least ambivalent: if it turns out that the past hero has been saved, this will be a particularly gratifying state of affairs in terms of value; she has been rewarded for her past heroic altruism. This person is *particularly* deserving of being saved. For me, it is only in this sort of case that, I feel, the act of SSA is probably not a mistake; that it can be defended.

If so, this, first, further supports my arguments concerning *the error of saving an ordinary person through SSA, because of Counting Oneself as a Hero*. Second, we also seem to have here some support for the even more paradoxical idea of safeguarding the *Past-Virtuous*. For, if even those who think that self-sacrifice for the sake of an ordinary person is a mistake, do not think so in a case of saving a past hero with a similar record, this matters. It suddenly makes it quite persuasive to hold the view that perhaps the past

¹⁴ Compare that there are many things that it would be morally wrong to do, but one can nevertheless be happy if they occur; see Saul Smilansky, "On Not Being Sorry about the Morally Bad" in *10 Moral Paradoxes*.

hero ought not to sacrifice herself so readily, for an ordinary person, because of her status as a person of unusual value (on account of her past SSA).

Once we focus on the fact that the loss of the self-sacrificer in typical cases of SSA is a special loss since the person is particularly worthy, we arrive at the puzzle concerning self-sacrifice I am raising here, and at the disturbing conclusion that our positive view of typical cases of SSA might well be mistaken.¹⁵

¹⁵ Versions of this paper were given at the Department of Philosophy, University of Turku, at the Department of Philosophy, University of Haifa, and in the Israeli Philosophical Association Annual Meeting; I am grateful to many participants for their comments. I am very grateful to Lucy Allais, Aliza Avraham, Aaron Ben-Zeev, Jonathan Berg, David Enoch, Amihud Gilead, David Heyd, Guy Kahane, Arnon Keren, Iddo Landau, Ran Lanzet, Sam Lebens, Arad Levin, Tal Manor, Jeff McMahan, Ariel Meirav, Juha Räikkä, Alma Smilansky-Teichner, Daniel Statman, Rivka Weinberg, and anonymous referees for the *Journal of Controversial Ideas*, for helpful comments on drafts of this paper.