The Worth of Moral Worth

Alisabeth Ayars

Draft as of 10/9/26

Invited paper, Aristotelian Society, 23 November 2026; to appear in *Proceedings of the Aristotelian Society (2027)*

Abstract. I investigate a neglected question about moral worth: what is valuable about acting with moral worth? I argue that acting with moral worth is valuable because it realizes a certain kind of teleological structure in a fragment of the universe: one in which morally significant features of reality make a difference to what happens through their normative significance. In the central case, the badness of suffering explains its prevention by functioning as the agent's reason for that response, rather than merely as a cause, and thus makes a difference to what happens through the agent's will. I call this a strongly moral-teleological structure. I then use this account of the value of moral worth to illuminate its conditions: worthy action requires that the relevant moral consideration govern the agent's response as the reason it is.

> A man said to the universe:
> "Sir, I exist!"
> "However," replied the universe,
> "The fact has not created in me
> A sense of obligation."
>
> Stephen Crane (1899)[1]

I. The Good Will

Today I was reading more about the atrocious practices associated with factory farming, of which I am already aware, but to which I keep returning as if I might read something that gives me a glimmer of hope. Let me tell you some facts I learned today. One of the most aversive experiences a human being can have is breathing in CO2 gas. The gas triggers extreme panic and a sense of impending doom. People report an overwhelming desire to escape. This is an ancient response; it survives even in people with damaged amygdalas.

The vast majority of farmed pigs are stunned using this method. They are put into a chamber that is lowered into a concentration of 90% CO2—far higher than the 35% needed to induce the panic

[1] Crane 1899.

response in humans.[2] They scream, they thrash, they throw themselves against the walls before losing consciousness after 30 to 60 seconds. It is true that the experience is brief; but I imagine that it feels much longer when you are the subject of it. And of course in the aggregate, the suffering is unfathomable. The Animal Welfare Institute estimates that roughly 117 million pigs undergo CO_2 slaughter each year in the United States. If each experienced an average of twenty seconds of conscious suffering, their suffering would together occupy more than seventy years—within a single year, in a single country.[3] Add other countries, and it multiplies.

Learning these facts this morning, I experienced something of a panic myself. I think it is odd that people are able to discount the suffering of animals so easily. I find it easy to imagine being an animal: to imagine being lowered into a gondola and inhaling the gas. For an experience so primitive, I suspect that it feels hardly different for a pig than for a human. So learning these facts is, to me, like learning of a practice of gas chambers to which millions of human beings are subject each year; and it is utterly horrifying.

But the point of this essay is not to discuss these facts. It is to discuss the nature of a kind of value that, to my mind, is distinctive: the value we refer to when we speak of an action's having moral worth. My panic at the suffering produced in me an overwhelming desire to act—to do *something*. The thought came over me suddenly that I ought to leave academia and devote my life to this problem. There could be nothing more important than solving it, just as, during the time of chattel slavery, there could be nothing more important than eliminating that evil. Upon arriving at this decision—upon forming the intention to act, at least to try to do something, which created the expectation that I *would* act—I felt some relief. But what was the nature of the relief?

It was certainly not the relief of feeling that my decision would make a difference to the problem; for surely it would not. Even if I managed to get a job at the most effective organization working on the issue, my efforts would scarcely register against the forces sustaining it. Nor was it the relief of being reassured that I was a "good person." I recognize how easily all of this could sound like smug self-satisfaction; you will have to trust me that it was neither. My distress during this episode really was outward-directed, at the atrocity itself and the occurrence of the response, not at my own moral quality. The relief (however minor) was produced by a slight shift in my perception of the moral landscape itself. The atrocity was still there, but now the fact of my response was there too. I felt relief at the very movement of my will: at the grip the suffering had on it. Kant thought that the good will is difficult to detect.[4] But I do think we can sometimes recognize it in ourselves and know it to be genuine. Moral horror is real. It is possible to want an

---

[2] Feinstein et al. (2013)

[3] The aggregate calculation uses the Animal Welfare Institute's estimate that 90% of approximately 130 million pigs slaughtered annually in the United States undergo CO_2 slaughter. Assuming an average of twenty seconds of conscious suffering per pig yields approximately seventy-four cumulative years annually.

[4] Kant 1785, 4:407.

atrocity to end without ulterior motive, and to want this enough to forgo other goods, including one's own comfort. What produced my relief—minor relief, against the unmitigated atrocity out there, but still something—was the recognition of such a motive in myself.

The fact of the grip on my will was not simply good for me, or something that made me good, but good in itself: a good aspect of the universe.[5] It would be just as good if someone else were gripped rather than me; the identity of the person makes no difference. So the goodness at hand is impersonal: it is a good thing that people's wills can be gripped by reasons of the relevant sort. (A universe in which this occurs is better than a universe that is ruthlessly mechanical.) This is the kind of value I want to understand. To manifest good will—to be gripped by what is good or right, and repelled by what is bad or evil—is a good thing, and it is good independently of whether it has any effect. It is also good independently of its "speaking well" of the person, or rendering her liable to praise or deserving of credit. (These things may be true, but they are downstream of the value I seek to characterize. Acting for moral reasons is to a person's credit, presumably, because it realizes a kind of value for which the person is in some sense responsible, the value being tied to the exercise of her will. But we can consider the value of a good will independently of the question of in what sense that value can be credited to the person, which is a deep question in itself.)

I want to start by examining the behavior of the value in question: how it weighs against other things, how it aggregates (if at all), and the kind of reasons it gives us. I am deliberately leaving aside, for now, the question of what the good will, or acting with moral worth, consists in, which is the question that has dominated contemporary metaethical discussion. (A central debate, for example, is whether it consists in being motivated by the right-making features or by rightness itself. Is it enough, to instantiate the value, that I truly desire the end of the pigs' suffering, in the way I desperately desire a swell career and good health? Or must there be a distinctively normative or moral element to my motivation that is not realized in all cases of desire?)[6] My hope is that clarifying the nature of the value might help us answer this question, since the conditions of acting with moral worth are the conditions under which this special kind of value is introduced into the world. The thing whose value is under discussion is, of course, the thing whose nature philosophers have sought to characterize since Kant, in discussions in which

[5] There is precedent for the idea that being for the good is itself intrinsically good. Hurka 2001 develops a "recursive" account on which loving what is good is itself intrinsically good, an idea he finds in Moore (1903, ch. VI); Adams 2006 characterizes virtue as excellence in being for the good. Unlike Hurka, I do not take this goodness to sum; see "The Universe Made Moral" below.

[6] For the view that moral worth requires motivation by the right-making features, see Arpaly 2002 and 2003 and Markovits 2010; for the view that it requires a distinctively moral element in the motivation—motivation by rightness itself, or moral knowledge—see Sliwa 2016 and Johnson King 2020. The Kantian version of the question begins with Herman 1981. I take up the question in Ayars 2026.

Huckleberry Finn, moral fetishists, and other characters have played a prominent role.[7] But I am not taking—and indeed refuse to take—Kant as my starting point, despite "shines like a jewel" and such.[8] I want to start afresh, and I will bring in Kant only once we have something on the table. Nor will I distinguish, at the start, between the value of good will itself (the fact of being gripped by the relevant considerations, resulting in certain kinds of desires) and the value of acting from good will (expressing these desires in action).

I have thus far assumed a consequentialist framework for conceiving of the value associated with moral worth: the exercise of good will is a good thing. A world in which there are instances of will being gripped by moral reasons is better than a world in which no will ever responds to moral reasons (perhaps because there are no rational wills at all, or all wills are evil). This consequentialist framing comes from my sense that, in responding to the moral horror of factory farming as I did, something good happened. Let us first bring this idea under scrutiny.

One thing I want to say at the outset is that whatever consequentialist goodness the exercise of good will manifests, it does not offset the original badness in the world to which that will is a response. That my will, or anyone's, is gripped by the evil of suffering is not nothing, but it is cold comfort in the face of the magnitude of the bad itself.

To bring out the value at stake, it is perhaps helpful to compare a world in which badness exists and there is no good will, with a world in which badness exists and there is good will. The former was, of course, our world for hundreds of millions of years, as evolution did its damage and created nervous systems that could feel pain, and the entire multitude of suffering went unwitnessed and uncared for. Is it a good thing that creatures now exist who can witness it and be moved to stop it without any motive of self-interest? It is unclear, of course, whether the appearance of human beings is a good thing from the point of view of overall suffering; but the question is whether it is good in this one respect. And I say that it is, and significantly so. Without it the world would only ever have been a machine, blindly pushed around by desire and blind forces. This is the surest fact about the value of good will, it seems to me: surer than any fact about the value of a particular instance, or about how the goodness of the instances aggregates.

This is not to say that the instances are not good independently of their contribution to the existence of good will as a whole; they are, as my mild relief at the movement of my own will attests. But at a certain level of granularity our intuitions break down. It is good that my will was gripped by the suffering this morning; would it be better if I had had a friend with me whose will was also gripped? Would it add goodness to the world if each day I apprehended the suffering anew, and felt the pull anew? Would it be better if I were somehow acquainted with the suffering

---

[7] Huck Finn enters the discussion with Bennett 1974 and is central to Arpaly 2003; the moral fetishist is Smith's (1994, pp. 71–6).

[8] Kant 1785, 4:394.

of each individual pig, and felt the weight of its particular suffering, rather than the suffering of the aggregate? Such questions about subdivisions and aggregates strike us as, in some way, absurd. (Consider also the immediate difficulties we would run into in adjudicating whether, and to what extent, we have "reason to promote" instances of good will in ourselves and others, the usual heuristic for assessing consequentialist goods.)[9] These complications cast doubt on the simple consequentialist reading of good will's value.

Let me set these questions aside for the moment. I believe we will get a better handle on them once we have a more complete account of why good will strikes us as valuable in the first place. Why does it soothe us, to some degree, that good will exists at all? What is clear is that our response to good will is, in some important sense, existential. It is good because of what it means: what it indicates about our position or status, and the position or status of creatures generally, in the universe. The word we reach for is *mattering*.[10] The fact that the will of others can be, and is, gripped by our suffering means that our suffering matters in a way that it could not if there were no good will. The consoling thought behind the solace I took in my own reaction is that it secures, in some inchoate sense, that the suffering of the pigs matters, however impotent that mattering is in removing the evil to which our energies ought to be directed.

We must be careful, for there is a risk of interpreting the thesis about mattering as trite. The natural first reading of my claim is that the pigs' suffering mattered *to me*, and that this realizes a kind of relational value. The hypothesis would then be that good will is valuable because there is value in our mattering to other people, and good will is just the expression of this valuable form of concern. The story would go on to appeal to the general value of being relationally embedded, which is good for human beings given our deeply social nature. (Applied to pigs, the story is fuzzier; but perhaps there is something "objectively" good even for an animal in mattering to someone else, however little it cares whether it does.)

But this hypothesis misses the boat. It is generally good for us, of course, that we matter to others, since we are relational beings with social needs; but this does not touch the deeper existential concern. The manifestation of good will does not merely introduce a relational good, mattering-to-someone; it introduces a deeper, non-relational kind of mattering into the universe (or so the thought goes). The capacity to be gripped by others' suffering in this distinctive way gives suffering a metaphysical status it would not otherwise have. It is the transition from a wholly indifferent universe to one that can, through the kind of willing in question, manifest the right kind of sensitivity to pain. When I felt consoled at my own reaction to the pigs' suffering, there was behind it a feeling of power: the feeling that I had made it the case that the suffering

---

[9] On the difference between promoting a value and honouring it, see Pettit 1989. Scanlon 1998, ch. 2, argues that to value something is not in general to see reason to promote it.

[10] Compare Nagel 1971 on the thought that, from the point of view of the universe, nothing we do matters; Kahane 2014 on our cosmic insignificance; and Wolf 2010 on meaning in life.

really did count—that I could stand against the universe's prior indifference, not merely by opposing it, but by making it otherwise.[11]

But of course the deep and difficult question is how someone's suffering's mattering to me (or being the sort of thing that some will or other might be gripped by) could make it matter in some deeper metaphysical sense. This seems like a fallacy, and a grandiose one at that. None of us has the authority, by our pattern of concern alone, to make it the case that anything matters "objectively." There is also the notable fact that we can be wrong about what matters. It could not literally be true that someone's being gripped by a consideration makes it objectively count, in the sense of giving it status as a reason, since people disagree about what matters, and the resulting "objective truths" would be incompatible. No sane metaethical account could identify a thing's mattering objectively with its mattering to some person or other. Whatever power good willing possesses to introduce a kind of status into the world, then, it is not the power to make normative facts true by our acts of willing alone.

I have suggested that good will—being gripped by considerations such as the suffering of others, such that one is moved to act on their basis—is intrinsically good. The goodness at hand is not merely attributive (it is not just that it makes agents into good agents);[12] its presence raises the world's value, and we should be glad of it. A world with good will in it is better not merely than a world in which wills exist and are turned only to the bad, but than a world in which there are no wills at all to be oriented.

It is not, however, good in the way that pleasure is good, as a quantity to be summed. It is dubious that more instances of good will are always better. The goodness of good will is analogous to the goodness of a correct response. Correctness is not a summable quantity, but each instance of good will fully realizes the goodness in question. In being glad we register that here, in this respect, things are as they ought to be. Compare true belief. It is good that true belief exists, and each true belief manifests correctness; and yet there is no reason to manufacture believers, and the reason to believe truly is not that true belief is good.[13] I will return to this comparison.

I have further suggested that the presence of good will in the universe allows suffering to matter in a way that it otherwise would not; and the relevant kind of mattering is not the trite relational

---

[11] Something like this feeling is, I believe, behind the elaborate rituals surrounding death. Death introduces a threat: that the person is a mere thing, a causal object, capable of being bashed and broken, with no more power against the forces of the machine than any physical matter. Our careful treatment of the body as not a mere object, and the forms of respect we pay to the dead as if to reaffirm their personhood, are intended to fend off this threat, in part by making it false.

[12] On attributive goodness, see Geach 1956.

[13] On truth as the standard of correctness for belief, see Velleman 2000 and Shah and Velleman 2005. Kelly 2003 argues that the reasons for belief are not instrumental reasons deriving from the value of true belief.

variety but a deeper, metaphysical kind. The question is in what sense the presence of good willing introduces such mattering into the universe.

II. Mattering and Difference-Making

Good willing does not create the normative facts: as just noted, the thesis that our patterns of concern make them obtain is absurd. But altering the normative facts is not the only existentially significant change the existence of good will could make. Return to the phenomenology with which this essay began: the felt sense in which my responsiveness to the pigs' suffering provided some ground for relief. Was this the feeling of having brought a reason into existence out of nothing? No. The pigs' suffering is apprehended as already bad, and as bad in virtue of its features, irrespective of how anyone responds to it.

The problem is, rather, that it is all too much of a reason already. The pigs' suffering calls out—indeed screams—for some kind of response; and the distinctive angst is the worry that this call might go unanswered. It is the threat of the universe-as-machine, of heartless, witless nature, that the presence of good will addresses. The indifference of the universe is a threat independent of the threat posed by the evil within it: for suffering to have no impact on the universe, for the universe to fail to respond to it correctly, is a way in which suffering can fail to matter. It is this threat, I submit, that the existence of good will goes some way toward undermining. The feeling of power—the feeling that I had made it the case that the suffering really did count—was the feeling that I had made its normative significance efficacious: that I had made it real by responding to it as it should be responded to.

Here too we must be careful. A natural way to put the thought just expressed is that good will is good because it functions as a bridge between the normative and the causal order. The pigs' suffering is a bad thing; and with respect to this bad thing a variety of responses are fitting,[14] such as attempts to eliminate or prevent it. Good will is valuable, on this picture, because it takes the bad thing as input and enacts the fitting responses as output, allowing suffering to make the difference in the world that it ought to make. But the picture is too simple. Mere difference-making, even when the difference made is the fitting response, is insufficient to instantiate the kind of goodness we find in good will.

To see this, suppose a brute causal law linked suffering with a fitting cosmic response. Imagine that among the fundamental laws of the universe there were one specifying that suffering above a certain duration or magnitude is disallowed. The law would ensure that such suffering never enters the world. (From our point of view, it would simply mean that we never encounter it; some ordinary accident would always intervene, as in Lewis's time-travel case, where each of Tim's attempts to kill his grandfather is foiled by a failure of nerve or a strange noise.)[15] The

[14] On fittingness, see McHugh and Way 2016.

[15] Lewis 1976.

suffering of the pigs could not occur in this world. The law would fix on suffering itself, rather than on a physical correlate whose regulation happened to prevent it. Yet it would remain blind to suffering's badness. It would simply ensure that suffering never crossed a threshold, just as nature limits the speed of light.

Is this a universe in which suffering really matters? In one sense the law undoubtedly realizes an element of how things ought to be; for it ought to be that extreme suffering is no part of our world, and the law guarantees precisely that. We would be fortunate to live in that world. But it is clear that in this world we do not have everything we want. Difference-making is a causal notion; it entails only that Y would have taken a different value had X taken a different value, with the rest of the world held fixed.[16] Such difference-making does not track suffering's badness. Even granting the naturalist view that badness is constituted by suffering, so that a law fixing on suffering fixes on badness, the law does not respond to suffering *as* bad. The fittingness of the response is something you and I see from outside, using our own sense that suffering is bad. The difference-making is, in Kantian terms, an accidental feature of that universe. The law might just as well have fixed on pleasure rather than suffering; and it could just as easily have guaranteed the persistence of extreme suffering rather than its elimination. (These modal claims are equivalent in force to claims such as "the speed of light could easily have been 200,000 kilometers per second rather than 300,000.")

To draw the analogy to moral worth, imagine an agent who takes the prevention of suffering as her non-instrumental end without responding to suffering as bad. (Perhaps she locks onto it arbitrarily, in the manner of a compulsion; or she responds to it as something else, as distasteful, say, but not morally significant.) This agent, like the law, makes suffering a genuine difference-maker: because she takes its prevention as her end, a bad thing now has causal consequences it would not otherwise have had, and the consequences are exactly those a fitting response would produce. But her response hardly strikes us as manifesting the correctness of the morally worthy agent's response.

Consider, then, a second possibility: a law, of the ordinary causal-nomological sort, which fixes on badness itself. The law says: whatever is bad does not occur. Such a law would track suffering's badness, and so seems to respond to suffering as bad. Would it get us the mattering we desire?

There is a real risk that such a law is unintelligible. The natural way for badness to produce a response is for someone to recognize it and respond to it as such; if we remove the someone, we seem to remove the thing that could make the connection make sense. But let us try. Under such a law, badness would be the property in virtue of which the law applies; it would occupy the same explanatory role that a nonnormative property, such as suffering, occupies in an ordinary causal law. Certain conceptions of the law of Karma are consistent with this picture: the law has

[16] Lewis 1973.

no administrator, and happiness and suffering arise as consequences of actions in an impersonal causal order.

At first glance, the law may seem to get us what we want, since suffering now not only systematically makes a difference but does so because it is bad. But this would still not get us a universe in which suffering really matters. Suffering now fails to occur because it is bad only in the sense in which a body falls because it has mass: badness is the condition the law specifies, and the law operates on it as it would on any other magnitude. The law is therefore as arbitrary as any other physical law. There is no particular reason why it connects badness to a fitting rather than an unfitting response. If the law is not selected for any reason, then a law saying "whatever is bad is guaranteed to occur"—the inverse of the fitting law—is just as available, and the fact that we got the good one is just an accident.

Another way to put the point is this. The pigs' suffering, I said, calls out for some kind of response, and the distinctive angst is the worry that this call might go unanswered. Has the call been answered in this universe? No. The badness of the pigs' suffering has not been responded to *as such*; the badness does not explain the response *in the right way*, for the response has nothing to do with the *normative significance* of badness. To say that the pigs' suffering calls out for a response is to say that it constitutes a reason for a response. An arbitrary causal link between badness and a fitting response does not heed this significance.

Let us consider a third possibility. Thus far we have considered laws which posit a brute causal relationship between suffering, or its badness, and fitting responses. One might suspect that the problem is that a causal connection is necessarily arbitrary—this is what makes the unfitting laws equally available—and that it can be fixed by invoking some form of non-causal explanatory link. In "Why Anything? Why This?",[17] Parfit considers the possibility that reality as a whole has a non-causal explanation: its existence or nature might be explained by the fact that it is simple, or beautiful, or good. Suppose the goodness of the universe explains its nature. This theory, the Axiarchic View, can be expressed, Parfit says, by the claims:

(1) It would be best if reality were a certain way.

(2) Reality is that way.

(3) (1) explains (2).

The nature of the "because" here is left unspecified, save that it is non-causal. (Parfit acknowledges that it "cannot be easily explained.") Does the Axiarchic View get us the kind of mattering we want? (To make the case parallel to the others, imagine that badness selects as well as goodness; the Axiarchic View then says that the non-existence of extreme suffering is explained by the fact that reality would be worse if it contained it.)

[17] Parfit 1998. "Axiarchism" is Leslie's term. The view is developed at length in Leslie 1979.

Whether it does depends entirely on what grounds the "because." *Why* does the badness of suffering explain its non-existence? One possibility is a mere nomological connection: the badness explains the response because a brute meta-law specifies badness as the relevant selector. This would be unsatisfactory. It would mean that the universe has a brute fact at the top, that badness selects; the law is accidental one level up.

Can we make sense of another possibility: that goodness and badness select through their normative significance alone? Suppose we say that it is not a brute fact that goodness rules. Rather, the good is *what ought to be*, and that is why it is; the bad is *what ought not to be*, and that is why it is not. This is the possibility that should interest us. If badness selects through its normative significance, its explanatory role is not arbitrary. The unfitting law is no longer equally available, since it would not be responsive to the significance of suffering's badness, and so would have no explanation. This is what it would take for suffering to really matter in the universe: for its normative significance to be explanatory in this way, so that the need for brute nomological connections is removed.

The question is what it is for normative significance to be explanatory in this way. As we have seen, it is not merely for the universe to be differentially responsive to goodness or badness; this could be achieved by brute laws that are not sensitive to what responses are called for. The natural hypothesis is this. What is it to say that badness is normatively significant? It is to say that badness gives reasons: reasons to disprefer it and to prevent it. So for badness to explain an outcome through its normative significance is for the *reason it gives* to explain the outcome: for the badness of suffering to explain suffering's elimination is for suffering to *function as a reason* for its elimination. This requires a more specific structure, which we might call *strongly* teleological:

(1) It would be best if reality were a certain way.

(2) Reality is that way.

(3) (1) explains (2) *by being, and functioning as, a reason for (2)*.

The interpretation of the third clause is important. To say that (1) explains (2) by functioning as a reason for (2) is not to say that there is an additional fact, that (1) is a reason for (2), which bears a brute explanatory relationship to the connection between (1) and (2). It is to describe the way in which (1) explains (2). A reality's goodness explains its existence by functioning as a reason for reality's being that way: the universe, in some sense, *takes* the goodness as a reason for bringing about that reality, and this grounds the reality's existence. This, I submit, is the valuable form of explanation: the explanatory structure that, when realized, is good. For suffering to matter in the sense we desire is for its badness to function as a reason: for it to be responded to *as the reason it is*, to be eliminated or prevented because that is what it *calls for*.

One question that arises is whether what is valuable is the realization of strong teleology as such, or its realization in the moral realm. I am inclined to think the latter: a strongly moral-teleological universe has a special kind of value, over and above that of a rational universe more generally. I will say more about this in the next section.

We have not yet said what role the good will plays in bringing about this valuable structure. Could the structure be instantiated without a will? Could suffering's badness function as a reason without anyone to appreciate it? I think not. Strong teleology, plausibly, requires at least a minimal subject. A reason explains an outcome only by being acted on (or treated as a rational basis); otherwise, it remains an inert normative fact. But to act on a reason is to take it as such—to register it and be moved by it. A universe that took suffering as a reason would itself be, or contain, a subject capable of governing its own activity.

The point can be made another way. An agent who relieves suffering because it is bad (or because it has the features that make it bad), and so acts with good will, has an orientation that can be evaluated as correct or incorrect. Her taking the suffering as a reason is correct if suffering is bad, and incorrect otherwise; the presence of a reason is the *standard* to which the attitude answers, and at which it aims. If the badness of suffering explains the outcome as a reason rather than as a cause, something in the explanation must be the kind of thing that gets the reason right. So if the universe took suffering as a reason, the universe, or some part of it, would have to be in a state with correctness conditions of this kind; and it is difficult to see how such a state could be anything but a state of a subject. Such a state is an attitude: it presents things as being some way, and it can be assessed for whether they are that way. For a reason to explain, it must be taken as such; for there to be a taking, there must be something capable of correctness; and only a subject can be correct.

But we need not settle the question here; none of the claims of this essay depend on it. Good will is, as far as we know, the only source of strong teleology in our universe. A will can be governed, not merely caused; it can grant a consideration authority, correctly or incorrectly; and so it can make it the case that suffering explains through its normative significance.[18] Whether or not a will is necessary to realize this structure, good will does realize it. Its special value lies in enabling considerations that are reasons to function as the reasons they are.

III. The Universe Made Moral

The theory we have been developing can be put as follows.

Teleological View of Moral Worth: Acting with moral worth is valuable because, when an agent acts with moral worth, she realizes a strongly moral-teleological structure. The moral badness of suffering (or the moral significance of some other consideration) explains the

[18] The conception of rational agency as self-government, on which to act for a reason is to grant a consideration authority over oneself, is Korsgaard's; see Korsgaard 1996 and 2009.

response by functioning as the reason for it: the agent correctly takes the moral badness as a reason, and she prevents the suffering because she so takes it. Suffering is then prevented because it *ought not to occur*. The "because" is the sort a reason supplies, rather than the sort a brute law supplies.

On the Teleological View, then, the conditions of moral worth—the conditions under which an agent acts with moral worth—are the conditions under which this valuable structure is realized.

What, exactly, is the explanandum in the teleological explanation, the fragment of reality that the normative significance explains? It is often assumed that acting with moral worth requires acting rightly: it is not enough that an agent feels something; the concern must issue in action, and the action must in fact be right.[19] We can now see why. An agent's right action provides what "ought to occur" in the teleological explanation, and to say that it ought to occur is not just to say that the agent should have done it; it is to attribute value to the response itself. It ought to be that suffering is prevented, that the needy are helped, that promises are kept, that creatures are not harmed; and it ought to be that the universe conspires to bring these things about. When an agent acts rightly, then, she makes a fragment of reality as it ought to be; when she acts rightly because she is gripped by the moral considerations at hand, the explanation runs through their normative significance, as strong teleology requires. And even where inability or interference prevents right action—as with the pigs, where there was little anyone could do—there is goodness in the gripping itself. My impulse to help the pigs was not mere feeling but a movement of the will, a motive that would have issued in action had action been available. This is what makes my gladness intelligible: I had recognized that the ingredients of strong moral teleology were present. I was reminded that the universe can be responsive.

The theory specifies that the value of acting with moral worth consists in its realizing a *moral*-teleological structure: in the moral features of the universe being given their due weight. This matters, since otherwise it would be a theory not of the distinctive value of moral worth but of rationality more generally. What is special about a morally responsive universe is that it responds to the value of the individuals within it. The pig's suffering matters because the pig matters. When an agent takes the badness of its suffering as a reason, she makes it the case not only that the suffering counts, but that the pig counts: its standing is acknowledged in the governance of her will. Here the explanation reaches a fundamental moral value. It is good that beings who matter in their own right should have their standing acknowledged—that their significance should find expression in the world's response to them.

We can now see why the value of moral worth is not straightforwardly amenable to consequentialist thinking. That the universe is responsive to moral value at all, through rational willing, is a good thing: such a universe is better than one that does not respond. Each instance of

[19] The assumption is built into Markovits's Coincident Reasons Thesis (2010) and into Arpaly's account (2002), both of which define moral worth for right actions.

morally worthy action realizes the teleological structure in a fragment of the universe, and so, in a sense, brings the universe closer to being as it should. But the value of the whole is not the sum of the values of its fragments. A universe in which rational willing answers to moral value is better than one in which it does not. But this is a difference in kind, not in magnitude; so multiplying instances of good will does not increase it. This is why it is natural to speak of the goodness that *inheres* in an instance of moral worth rather than the goodness it *adds.*

It also explains why talk of "reasons to promote" good will is inapt. To *aim* at producing worthy actions is to treat them as units of this kind, and this aim is self-undermining. Worthy action is responsiveness to moral reasons, and an action performed to increase responsiveness in the world responds to the value of responding rather than to the reasons that make responding worthy. Consider again the analogy with true belief. Though true beliefs are good, the reason to believe truly is not that true belief is good. The value is realized only by an attitude that is answerable to something other than the value itself: the first-order reasons for and against responses.

Let me close by clarifying what distinguishes the Teleological View from its alternatives. It contrasts, first, with views on which moral worth manifests only attributive goodness: the goodness of the agent herself.[20] On these views, the agent's goodness is fundamental; the action is good only as a specimen of its kind, much as a well-made knife is good as a knife. The Teleological View reverses this order of explanation. It allows that acting with moral worth merits admiration and credit, but grounds that credit in the distinctive value the action realizes. The agent deserves credit because this realization is her doing: it is she who takes the consideration as a reason. What matters on the Teleological View is that considerations come to govern a response as the reasons they are. The Teleological View is also opposed to views that acknowledge a value in acting with moral worth but situate it in something else, such as the relational value of mattering to someone.

This suggests a corresponding conception of what the concept of moral worth is for. Its function is not, fundamentally, to distinguish good agents from bad, or to regulate our practices of praise and admiration . Rather, it is to identify where this distinctively valuable moral-teleological structure is realized: where moral reasons come to govern what happens.

IV. The Conditions of Moral Worth

What does the Teleological View buy us? What do we gain, in our thinking about moral worth, from knowing what its worth consists in? I said at the outset that I would leave aside what acting with moral worth consists in, hoping that an account of its value would help. We are now in a position to see how. The Teleological View gives us a criterion: the conditions of moral worth are the conditions under which the valuable teleological structure is realized. This requires, at least,

[20] The literature standardly frames moral worth in terms of praiseworthiness or credit; see Lord 2017 and Singh 2020.

that the relevant moral consideration—the pig's suffering, or its badness—*function as a reason* for the agent. The agent must *take* it as a reason, and her action must be explained by this taking.[21] If the connection between suffering and the agent's response is not due to this taking, then it is like the blind law, however reliable it may be.

Earlier, I considered whether desire alone is enough to realize this structure. Is it enough for good will that I non-instrumentally desire the cessation of the pigs' suffering, in the way I desire a good career and good health, and act on this desire? The answer depends on what we take desire to be. If desiring the end of their suffering is a way of responding to their suffering as bad, then it is enough; desire, in that case, just is good will. But a desire that merely takes an end as its object and moves the agent toward it, without the agent taking suffering to provide a binding reason for the response, stands to suffering as the blind law stood to it. The Humean might insist that to desire an end is already to see it in a favorable light, so that anyone who acts on a desire for the suffering to end already takes its removal to constitute a reason.[22] Perhaps so; but then the seeing does the work in explaining the value of the response.

Discussions of moral worth often center on the concept of accidentality: to act with moral worth requires that one's motive bear a non-accidental connection to the act's rightness.[23] Here the Teleological View is, I believe, especially illuminating. To say that a right act is non-accidental just is to say that the teleological explanation is genuine: the action really is explained in the right way by its normative significance. Accidental right actions are ones in which the explanation is disrupted. The explanation may run through the wrong feature: the agent is not differentially responsive to the relevant consideration at all, but to something that happens to correlate with it, as Kant's shopkeeper is responsive to his advantage rather than to honesty. Or it may run through the right feature in the wrong way: the consideration moves the agent, but not as a reason, as when someone merely desires something without seeing it as providing a reason. We can also see why mere reliability is not enough for non-accidentality. Many philosophers have noted that the good-willed agent not only does the right thing here and now, but would do the right thing in nearby cases as well.[24] But a blind law that fixes on suffering is more reliable than any agent could be, and it does not diminish the universe's indifference.

Exactly what conditions the agent's will must satisfy to realize the teleological structure is beyond the scope of this essay; but the Teleological View at least provides a direction for inquiry.

---

[21] I defend a related requirement in Ayars 2026.

[22] For desire as a way of seeing its object as good, see Stampe 1987; for doubts about the guise of the good, Velleman 1992.

[23] See Herman 1981, Johnson King 2020, and Singh 2020.

[24] Herman 1981 is the classic statement of the point, made against the agent who acts from sympathy; see also Sliwa 2016 and Lord 2017.

The fundamental question is not what makes an agent praiseworthy for acting rightly, but under what conditions her right action manifests the right kind of response to reasons.

I will conclude by returning to the phenomenological observation with which this essay began. What was the nature of the relief I experienced? It was not relief that the pigs would be helped (they won't), and it was not relief that I was a good person (anyone's responsiveness would do). It was relief that indifference was not the whole truth about the universe. For a moment, in one small part of it, another creature's existence and suffering did create a sense of obligation. The power I felt was the power to make Crane's answer false in one corner of reality.

The demand made by Crane's man is, in one sense, unintelligible: nature, understood as the mere unfolding of impersonal laws—the witless, churning machine—cannot acknowledge an obligation. But the demand is, in another sense, met. The significance of suffering is intelligible to a being for whom another's existence can make a claim upon the will—that is, to us. And we, too, belong to the universe. In our responsiveness, one part of reality answers to another; a relation obtains that is not just causes-to-happen, but is-a-reason-for. And the emergence of this relation from a world of mere happenings is, in a certain sense, miraculous. Within good will, therefore, lies the power to make suffering matter, by responding to it as it should be responded to. Its badness can find acknowledgment within the very world that permits it. The existence of good will gives us the assurance that when suffering cries out for a response, its cry is not, everywhere and always, in vain.

References

Adams, Robert Merrihew 2006: *A Theory of Virtue: Excellence in Being for the Good*. Oxford: Clarendon Press.

Arpaly, Nomy 2002: 'Moral Worth'. *Journal of Philosophy*, 99(5), pp. 223–45.

Arpaly, Nomy 2003: *Unprincipled Virtue: An Inquiry into Moral Agency*. New York: Oxford University Press.

Ayars, Alisabeth 2026: 'Moral Worth beyond Right Reasons'. In Russ Shafer-Landau (ed.), *Oxford Studies in Metaethics*, vol. 21. Oxford: Oxford University Press.

Bennett, Jonathan 1974: 'The Conscience of Huckleberry Finn'. *Philosophy*, 49(188), pp. 123–34.

Crane, Stephen 1899: *War is Kind*. New York: Frederick A. Stokes.

Feinstein, Justin S., et al. 'Fear and panic in humans with bilateral amygdala damage.' Nature neuroscience 16.3 (2013): 270-272.

Geach, Peter T. 1956: 'Good and Evil'. *Analysis*, 17(2), pp. 33–42.

Herman, Barbara 1981: ‘On the Value of Acting from the Motive of Duty’. *Philosophical Review*, 90(3), pp. 359–82.

Hurka, Thomas 2001: *Virtue, Vice, and Value*. New York: Oxford University Press.

Johnson King, Zoë 2020: ‘Accidentally Doing the Right Thing’. *Philosophy and Phenomenological Research*, 100(1), pp. 186–206.

Kahane, Guy 2014: ‘Our Cosmic Insignificance’. *Noûs*, 48(4), pp. 745–72.

Kant, Immanuel 1785: *Groundwork of the Metaphysics of Morals*. Translated by Mary Gregor. Cambridge: Cambridge University Press, 1997. Cited by Academy volume and page.

Kelly, Thomas 2003: ‘Epistemic Rationality as Instrumental Rationality: A Critique’. *Philosophy and Phenomenological Research*, 66(3), pp. 612–40.

Korsgaard, Christine M. 1996: *The Sources of Normativity*. Cambridge: Cambridge University Press.

Korsgaard, Christine M. 2009: *Self-Constitution: Agency, Identity, and Integrity*. Oxford: Oxford University Press.

Leslie, John 1979: *Value and Existence*. Oxford: Basil Blackwell.

Lewis, David 1973: ‘Causation’. *Journal of Philosophy*, 70(17), pp. 556–67.

Lewis, David 1976: ‘The Paradoxes of Time Travel’. *American Philosophical Quarterly*, 13(2), pp. 145–52.

Lord, Errol 2017: ‘On the Intellectual Conditions for Responsibility: Acting for the Right Reasons, Conceptualization, and Credit’. *Philosophy and Phenomenological Research*, 95(2), pp. 436–64.

Markovits, Julia 2010: ‘Acting for the Right Reasons’. *Philosophical Review*, 119(2), pp. 201–42.

McHugh, Conor and Jonathan Way 2016: ‘Fittingness First’. *Ethics*, 126(3), pp. 575–606.

Moore, G. E. 1903: *Principia Ethica*. Cambridge: Cambridge University Press.

Nagel, Thomas 1971: ‘The Absurd’. *Journal of Philosophy*, 68(20), pp. 716–27.

Parfit, Derek 1998: ‘Why Anything? Why This?’. *London Review of Books*, 20(2), 22 January, and 20(3), 5 February.

Pettit, Philip 1989: ‘Consequentialism and Respect for Persons’. *Ethics*, 100(1), pp. 116–26.

Scanlon, T. M. 1998: *What We Owe to Each Other*. Cambridge, MA: Harvard University Press.

Shah, Nishi and J. David Velleman 2005: 'Doxastic Deliberation'. *Philosophical Review*, 114(4), pp. 497–534.

Singh, Keshav 2020: 'Moral Worth, Credit, and Non-Accidentality'. In Mark Timmons (ed.), *Oxford Studies in Normative Ethics*, vol. 10. Oxford: Oxford University Press, pp. 156–81.

Sliwa, Paulina 2016: 'Moral Worth and Moral Knowledge'. *Philosophy and Phenomenological Research*, 93(2), pp. 393–418.

Smith, Michael 1994: *The Moral Problem*. Oxford: Blackwell.

Stampe, Dennis W. 1987: 'The Authority of Desire'. *Philosophical Review*, 96(3), pp. 335–81.

Velleman, J. David 1992: 'The Guise of the Good'. *Noûs*, 26(1), pp. 3–26.

Velleman, J. David 2000: 'On the Aim of Belief'. In his *The Possibility of Practical Reason*. Oxford: Clarendon Press, pp. 244–81.

Wolf, Susan 2010: *Meaning in Life and Why It Matters*. Princeton: Princeton University Press.