

The Actual and the Expected

Kurt Gerstein ordered and distributed a chemical substance for drinking water purification and disinfection at the German front in World War II. When he learned in 1942 that concentration camps wanted to use the substance – prussic acid, more infamously known as Zyklon B – to exterminate Jews in mass numbers, he tried to stop the killings. At considerable risk he sought to warn the broader world about the exterminations through various church organizations, a Swedish diplomat, and a Dutch underground group, and he sabotaged some shipments of Zyklon B. All of these attempts failed to stop the killings.¹ One characterization of Gerstein has him diluting a Zyklon B shipment in order to render it unusable – but the camps employed it anyway, resulting in slower, more painful and horrific deaths for the victims.² I think we should say that Gerstein did the wrong thing for the right reason. And I think it is important to preserve both elements of that evaluation, both the failure and the success. This paper is a defense of the necessity of two distinct – and in the more interesting cases, conflicting – moral evaluations of cases like Gerstein’s.

A dominating distinction in the case is that between actual outcomes and expected outcomes. If we judge the case by the actual effects of Gerstein’s choices, the evaluation is negative: Gerstein really ended up doing nothing to help suffering and dying concentration camp victims. But if we judge the case by the expected effects of Gerstein’s choices, the evaluation is positive: it was reasonable for Gerstein to think that *something* he did would reduce the number of killings.

¹ Saul Friedlander, *Kurt Gerstein: The Ambiguity of the Good* (New York: Alfred A. Knopf, 1969); and Pierre Joffroy, *A Spy for God: The Ordeal of Kurt Gerstein* (New York: Harcourt Brace Jovanovich, 1970).

² The dilution of the prussic acid and the attendant increased suffering is depicted in Costa-Gavras’s film dramatization of Gerstein’s life, *Amen*. (2002). The dilution and extra suffering is not mentioned in Gerstein’s own report of his activities during the war and may well not be historical. With this caveat, I will refer to the dilution and extra suffering several times in this paper.

The distinction between actual and expected outcomes is of course common among consequentialists.³ Less commonly recognized is that non-consequentialists also must take a side here too. Outcomes count for something in almost all ethical theories.⁴ Any deontological theory, for instance, that includes even a weak positive obligation to aid others must say whether fulfilling that obligation tracks what one reasonably expects to happen or what actually happens.⁵ This can't be emphasized enough: the issue is crucially underdiscussed by Kantians and deontologists.

I. The Case for Actual Outcomes

It's hard to celebrate Gerstein when nothing he did seems to have minimized suffering and death. Right action should be characterized by victims and what happens to them, and Gerstein helped few or no victims.⁶ That is a first argument in favor of evaluating by actual outcomes.

A second argument is this. The key facts in everything from science to athletic accomplishment are out there in the world, outside the agent – credit goes to the scientist who *in*

³ See Henry Sidgwick, *The Methods of Ethics*, 7th edition (Indianapolis: Hackett, 1907/1981), pp. 206-208, 413; J. J. C. Smart and Bernard Williams, *Utilitarianism: For and Against* (Cambridge: Cambridge University Press, 1973), pp. 46-56; Peter Railton, "Alienation, Consequentialism, and the Demands of Morality," *Philosophy and Public Affairs* 13:2 (Spring 1984): 134-171; David Brink, *Moral Realism and the Foundations of Ethics* (New York: Cambridge University Press, 1993) pp. 257-258; and Frank Jackson, "Decision-theoretic Consequentialism and the Nearest and Dearest Objection," *Ethics* 101 (April 1991): 461-482. The distinction even figures in importantly but without fanfare in Peter Singer's well-known "Famine, Affluence, and Morality," *Philosophy and Public Affairs* 1:3 (Spring 1972): 234.

⁴ Shelly Kagan, *Normative Ethics* (Boulder, CO: Westview Press, 1998), pp. 26, 66.

⁵ For a non-consequentialist defense of evaluating by expected outcomes, see Alan Donagan, *The Theory of Morality* (Chicago: University of Chicago Press, 1977), and Barbara Herman, "What Happens to the Consequences?" in Herman, *The Practice of Moral Judgment* (Harvard: Harvard University Press, 1993), pp. 94-112. For a non-consequentialist position that sometimes relies on actual outcomes – particularly on what happens to victims of certain actions – see Frances Kamm, *Morality, Mortality: Volume II* (New York: Oxford University Press, 1996) and *Intricate Ethics* (New York: Oxford University Press, 2006).

⁶ Or at least it's hard to argue that he did. On the one hand, had Gerstein resigned, perhaps his SS replacement would have sent more Zyklon B to the camps, or would have been more aggressive in looking for an even more lethal substance. On the other hand, his replacement wouldn't have tried dilution, and that would have subtracted some of the suffering that Gerstein unintentionally increased; see the following paragraphs in the main text.

fact discovers a cure for Huntington's disease, not one who merely tries; and credit goes to Abdul-Jabbar for actually scoring 38,387 points, not for merely being on the court. Ethics should also track facts out in the world. Gerstein in fact accomplished nothing for his intended beneficiaries.

A third argument concerns regret. If only expected outcomes matter, then Gerstein did the right thing, and he would have no reason to regret what he did. But of course it's reasonable for Gerstein to feel enormous regret over not acting differently. Any number of other actions could have in fact done more to stop or slow the genocide.⁷ Regret only makes sense if actual outcomes matter in moral evaluation.

Fourth, consider cases where two misinformed people fight. Each plausibly (but mistakenly) thinks that the other maliciously intends harm. Suppose Anderson and Bradley are looking for their lost children in the basement of the high school: it has been announced at halftime that a local killer has been spotted in the building. Anderson encounters Bradley in the dark basement, and Bradley looks like the description of the killer. Bradley encounters Anderson in the same dark basement, and Anderson similarly matches the description. If expected outcomes matter, both are justified in fighting one another.⁸ But it seems somewhat strange to say this – that both do the right thing. The fan of actual outcomes need not say that both do the right thing.

Next, suppose Herz knows something that Gerstein does not – something that could make Gerstein's actions yield better results. For instance, suppose Gerstein dilutes a shipment of

⁷ Friedlander argues that Gerstein would have been more successful had he worked with religious leaders inside and outside the Confessional Church to protest the exterminations. Public protests by German religious leaders had stopped an earlier National Socialist euthanasia campaign against the mentally handicapped (Friedlander, pp. 76-78, 136-137). But Gerstein wasn't well-informed on what had been successful against this campaign and why. And Gerstein couldn't have expected the Allies and neutral powers to do absolutely nothing about the exterminations, and he couldn't have been expected to know that they already had most of the information he tried to pass on. In short, other actions would have had better results, but he wasn't in a position to know this.

⁸ See Charles Friend, *Right and Wrong* (Cambridge, MA: Harvard University Press, 1979), p. 48.

Zyklon B with the reasonable expectation that it will be unusable at the camps.⁹ Suppose Herz also wants to stop the genocide; he has seen SS guards already use diluted Zyklon B in the camps and knows that it actually makes the victims (and no fewer of them) die more slowly and painfully. Herz knows of Gerstein's aims and is in a position to correct Gerstein's reasonable but incorrect beliefs. Herz should tell Gerstein that Gerstein *ought not* to dilute, and this is a moral "ought" – it is grounded in the actual facts that diluted Zyklon B will make the victims suffer more. But if only expected outcomes matter, Herz can't say that Gerstein morally *ought not* to dilute. If only expected outcomes matter, Gerstein's dilution sabotage is right, and can't be made more right with new information about the actual outcomes. This seems incorrect. Herz should tell Gerstein, and the best explanation for this is that morally it matters whether victims actually suffer more.¹⁰

There are important objections to using actual outcomes in moral evaluations. For instance, agents rarely could know in advance whether they were doing the right thing – moral evaluation might have to be retrospective. In fact the problem is even deeper: whether an agent has done the right thing wouldn't be fully settled until the end of the universe.

II. The Case for Expected Outcomes

Expected outcomes matter because Gerstein's attempts to sabotage Zyklon B shipments and expose the genocide were risky and brave.¹¹ Although he was unsuccessful, he seems

⁹ Gerstein was trained as a mining engineer, not a chemist. He did take steps to determine what would contaminate Zyklon B or cause it to decompose (see Friedlander pp. 186-188, 189-191), but his chemical expertise had its limits. (See also footnote 2 above. Again, while Gerstein seems to have sabotaged shipments of Zyklon B, the Costa-Gavras film's mention of related increased victim suffering is not historically well-grounded.)

¹⁰ Walter Sinnott-Armstrong, "Gert Contra Consequentialism", in *Rationality, Rules, and Ideals; Critical Essays on Bernard Gert's Moral Theory*, Walter Sinnott-Armstrong and Robert Audi eds. (Lanham, Md.: Rowman and Littlefield, 2003), pp. 145-163.

¹¹ This is the standard account of Gerstein. It is possible – though quite unlikely in light of his opposition to the Nazi regime earlier in his life, and the many documented personal risks he took during the war – that he

praiseworthy for trying. Vast numbers of his contemporaries were not so committed or courageous. We should commend Gerstein morally because of what he earnestly tried to do.

A second argument for expected outcomes relies on the broadly shared principle “ought implies can.” Morality’s obligations are tied to what it is possible for agents to know and do. Gerstein cannot have had an obligation to do something other than what he reasonably expected, on the best available evidence,¹² would be best. It is not fair to fault Gerstein for outcomes that he cannot reasonably have predicted.

A central point of morality is to guide decision making, a third argument notes. Since actual outcomes can’t be accurately predicted, they cannot guide decision making; only reasonably expected outcomes are relevant to decision guidance. To give up evaluating by expected outcomes is to give up morality’s key guidance role.

Fourth, if only actual outcomes matter, too much of morality is left open to luck. Gerstein was one of the earliest people to learn about the SS’s intention to utilize Zyklon B for mass killings. Once Gerstein has passed this information to various people – the Swedish diplomat he meets on a train, people working in a Dutch underground organization, and various sympathetic religious leaders – the rest of what happens is largely out of his control. What these people do with the information is up to them and to circumstance, and morally irrelevant to what *Gerstein* did, which is to fulfill a moral obligation to report and warn. Insulating moral evaluation against luck requires evaluating in light of expected outcomes.

misrepresented his efforts to stop the killings. What is both far more probable, and still correspondingly awkward for a Gerstein defender, is that he stayed in a role where he continued to order and distribute Zyklon B. He can’t have sabotaged all of these shipments. Gerstein seems to have both had faith that he would be able to do more good if he kept his post, and also a sense that his historical role was to provide testimony about the extermination’s perpetrators after what he saw as an inevitable defeat for Germany. See the Friedlander and Joffroy biographies.

¹² I think this is the most plausible version of an appeal to expected outcomes: expected outcomes are those that track the best available evidence. Other versions cash out expected outcomes in terms of what the agent herself believes, or perhaps in terms of what a population tends to believe at a given time. See Kagan, *Normative Ethics*, p. 65.

The previous section makes it easy to predict some objections to moral evaluations based on expected outcomes: evaluating by expected outcomes leaves out the harm to the victims and Gerstein's failure to help them; Gerstein's regret is unexplained; too many different (and perhaps conflicting) actions would turn out to be morally right; and more information from better informed third parties could not make Gerstein's actions morally any better.

III. Reduction

Most who recognize the distinction try to employ both kinds of evaluation, usually by making one reducible to the other. The hope in both approaches is to retain both actual and expected outcomes –one as central, the other on the cheap.

Consider first Jack Smart, who thought that actual outcomes were the bedrock of moral evaluation, with expected outcomes playing an important but reducible supporting role. In Smart's classic example, a person in 1938 who rescues a drowning man – a man who turns out to be Hitler – does the wrong thing, because the actual outcomes of saving Hitler are monstrous. But Smart adds that since the likely outcomes of most rescues are good, the rescuer should in fact be praised: a positive moral evaluation increases the likelihood of future rescues, by this person and by the audience for the moral admiration. Praise and blame are themselves secondary actions on this account, and they should track expected outcomes. Wrong actions can be praiseworthy, because praise itself is a separate, second-order action that can lead to better outcomes in future cases. On Smart's view, expected-outcome-based praise and blame are instrumental to more fundamental actual-outcome-based evaluations. So Smart's

consequentialism has a job description for expected outcomes – but it’s only a service industry job for actual outcomes, the real boss.¹³

Smart’s position makes sense of the regret someone like Gerstein should feel at the outcomes of his choices (and those choices themselves), and it’s plausible to think that praising people like Gerstein for taking personal risks aimed at beneficence will usually make the world better. But Smart leaves expected outcomes too much on the sidelines. First, Smart is stuck with the entailment that there is nothing intrinsically morally positive about Gerstein’s attempts to aid; Smart’s only reason for praising them is the bet that instrumentally similar attempts will pay off somewhere later or somewhere else. One way to put the point is this: Smart conflates praise and blame with *praiseworthiness* and *blameworthiness*. We want to know not just whether it’s a strategic good bet to praise Gerstein; we want to know whether he is in fact *worthy* of praise – whether his motives and intentions are aimed at the good. Smart could just as well praise a pathologically klutzy malicious sadist, someone so inept at inflicting harm that the more we praise him, the harder he tries and the more accidental good he does.¹⁴ Such people rarely exist, but the point is that Smart’s approach can’t distinguish them from Gerstein.

Other objections track the arguments of the previous section: Smart’s position violates the intuition behind the “ought implies can” principle, because Gerstein cannot reasonably perform the optimal action; Smart’s approach offers little by way of decision guidance; and he surrenders much of morality to luck.

¹³ J. J. C. Smart and Bernard Williams, *Utilitarianism: For and Against* (Cambridge: Cambridge University Press, 1973), pp. 46-56. Mill proposes a similar view in a footnote in the second chapter of his *Utilitarianism*.

¹⁴ Recall Monty Python’s Spanish Inquisition skit here. Spurred on by our praise, they would intend harm but in fact provide comfy pillows to increasing numbers of uncomfortable elderly people.

One can instead make expected outcomes play the lead role and actual outcomes the supporting role. A strategy contrasting deeply with Smart's is that of Barbara Herman.¹⁵ For Herman, the central object of moral evaluation is agent willings – that is, what the agent takes herself to be trying to do and why. Moral evaluation on this view focuses on what agents were trying to bring about and their reasons for trying. But Herman – and she is somewhat rare among Kantians for writing about the issue – does want some role for actual consequences. Suppose someone makes a promise to repay a loan that circumstances prevent her from keeping, say because an identity theft liquidates her bank account. On reasonable expectations, she took herself to be in a position to keep the promise, and that for Herman is central. But Herman thinks it also matters that the loan was actually not repaid. A person who is serious about keeping promises will respond to the failure to repay in some appropriate way, such as apologizing and arranging a new loan. And a person who is serious about keeping repayment promises should also guard herself against identity theft. In Kantian terms, Herman thinks the maxim to keep repayment promises should be supplemented with “maxims of response” (agent willings that respond to unexpected bad actual outcomes) and also “maxims of preparation” (agent willings that attempt to forestall bad actual outcomes). For Herman, moral evaluation is rarely evaluation of an unconnected discrete action; rather, morality also looks at the set of actions that precede and follow the act.

Herman's attempt to find room for actual outcomes in a Kantian framework – if only secondary room – is resourceful. On her view, the right thing for Gerstein to do in the face of failure is to form new “maxims of response,” maxims to try again to alert the world and help the suffering. Still, there is something left over that she cannot explain. Anyone in Gerstein's position would agonize and feel regret after his failures; but Herman says that “if there is no fault

¹⁵ Herman, *op. cit.*

in what the agent wills (in this wider context of assessment), it is not clear that there is anything that could count as a moral fault in what she does.” She claims that her view does have some room for agents to “regret that things did not work out,”¹⁶ but does not explain how her particular Kantian machinery can say this. And even if it somehow can, there is still something left out of the account: Gerstein would reasonably regret not just the actual outcomes, but that – particularly in the case of increased suffering from the diluted Zyklon B – it was *his actions* that led to them. Gerstein will regret not just what happens, but his own causal role in what happens. Herman’s attention to “maxims of preparation” doesn’t help – it’s not clear that Gerstein could have prepared in such a way that his outcomes would have been better. In sum, it’s not clear that Herman’s position (and others like it) can explain agent regret and the plausible and admirable array of *targets* of regret.

There are still other problems with approaches like Herman’s.¹⁷ Positions that focus on the intentions, expectations, and “willings” of agents – even an expanded set of willings like Herman’s – say too little about victims and what happens to them. Also, too little is said about sheer agent accomplishment and failure. Next, if someone had, with better knowledge and good intentions and willings, worked against Gerstein’s attempted dilution of the Zyklon B, Herman’s account would strangely leave both as doing the right thing. And the person with better knowledge couldn’t make Gerstein do something morally more right by informing him of the dangers of dilution.

¹⁶ Both quotes are from Herman, p. 104.

¹⁷ Less relevant to this paper, but still important, is that Herman doesn’t entirely succeed in avoiding agent evaluation (her other battlefield) I find her arguments on this issue strained.

IV. Morality's Two Subjects

The costs of reducing either sort of evaluation to the other are clearly too high. There remains the solution I would like to promote. Morality isn't about either actual outcomes or expected outcomes alone. Morality in fact has two subject matters. Morality is about two importantly different and irreducible topics. Call this the Two Irreducible Subjects Thesis (TIST).

I will characterize the dimension associated with actual outcomes as about *rightness and wrongness* and the dimension associated with expected outcomes as about *praiseworthiness and blameworthiness*. It matters less for now how we specifically characterize the two subject matters and more that they are distinct.¹⁸ The two dimensions usually line up: the right thing is usually praiseworthy, the wrong thing blameworthy. But they can also come apart, as they do in the case of Gerstein.

The distinction is ordinary and natural; that itself is an argument for it. We say of someone, "he did the right thing for the wrong reason," or "he did the wrong thing for the right reason." The latter is what we should say about Gerstein. Doing so preserves what is fundamental to the two kinds of evaluation. On the one side, Gerstein does the wrong thing: there are victims, whose suffering (and his role in that suffering) are clearly worthy of regret; and information from someone better informed could have helped Gerstein act rightly. On the other side, Gerstein is *praiseworthy*, and not just instrumentally to be praised: he takes courageous risks, does what is possible for him to do and what is possible for morality to guide him to do,

¹⁸ Here is an example of the sort of further specificity that might interest us. Evaluations of praiseworthiness and blameworthiness might take as objects motives, intentions, character traits, willings (Herman), or more generally the agents that have these. Evaluations of rightness/wrongness might take as objects actions, principles, rules, or consequences. *Actions* may actually be disputed territory, since *both* types of evaluation might naturally take *actions* as objects. We could characterize Gerstein's dilution of the gas as a beneficent action intended to help (and so praiseworthy) or a harmful action that increases victims' suffering (and so wrong).

and his praiseworthiness cannot so easily be disrupted by luck. Both evaluations are crucial if we are to capture the moral truth about Gerstein without a remainder.

Another argument for TIST is that morality has more than one point and purpose.¹⁹ Morality is a criterion of rightness; morality guides our decision-making; and morality determines the proper distribution of benefits and burdens. It's unlikely that all of these roles can be performed by one kind of evaluation.

Third, thinking of morality as two distinct subject matters yields rich solutions and explanations in importantly debated areas in normative ethics: the nature and plausibility of the intend/foresee distinction,²⁰ moral dilemmas, the problem of dirty hands, permissions to harm in self-defense, and the shape and scope of moderate thresholds in deontological ethics. TIST is supported not only by the consideration of actual and expected outcomes in this paper, but by its handling of these other issues as well.

For some the most important objection to TIST is that this bifurcation of moral evaluation threatens the moral coherence and meaning of our lives. Conflict between the two dimensions will usually be uncommon, but that won't satisfy the objector.²¹ A TIST defender should bite the bullet here. The moral world is a dangerous place, and the poor choices of others, and perhaps even natural conditions, can create circumstances where there is no way for certain agents to be both praiseworthy and right.²²

¹⁹ The language about morality's "point and purpose" is Shelly Kagan's (*Normative Ethics* pp. 6, 66).

²⁰ For instance, most attacks on the intend/foresee distinction fail to note that the permissions and requirements entailed by the distinction are not a function of a particular agent's occurrent intentions. See my "Morality's Two Subjects in the Intend/Foresee Distinction" and "Dirty Hands and Morality's Two Subjects," in manuscript. See also William FitzPatrick, "Acts, Intentions, and Moral Permissibility: In Defence of the Doctrine of Double Effect," *Analysis* 63:4 (October 2003): 317-321.

²¹ The frequency of conflict will be in part a function of social conditions. Saul Friedlander points out that totalitarian regimes, for instance, will generate situations like (if not in their drama and scale, at least in structure) Gerstein's. See Friedlander, pp. 226-228.

²² Ruth Barcan Marcus replies in a similar way about the problem of moral dilemmas in "Moral Dilemmas and Consistency," *The Journal of Philosophy* 77 (1980): 121-136.

What follows from TIST is significant. It means that much of contemporary moral philosophy harbors a crucial conflation. But much more is gained from thinking of morality as two matters instead of one.