RAIS Journal for Social Sciences | Vol. 10, No. 2, 2026
ISSN 2574-0245 (Print) | ISSN 2574-1179 (Online) | DOI:10.5281/zenodo.23097971

Is It Ever Wrong to Do the Right Thing for the Wrong Reasons?

Max Reiter

Delbarton School, mjreiter08@gmail.com

Abstract: This paper examines whether an outwardly right deed can nevertheless be wrong when performed for a morally objectionable reason. It argues that the apparent paradox arises because moral judgment can target two aspects of one intentional action: the deed, considered through its public effect, and the will, considered through the operative reason that helps constitute the act. Drawing on Anscombe and Davidson, the paper treats motives not as external additions to behavior but as part of the action's intentional description. The paper develops a conventionalist account of moral standards grounded in a community's collective endorsement, distinguishing settled norms from temporary majority opinion. Within this framework, a deed such as stopping an attack may remain right even when the fuller action—stopping it from racial hatred—is wrong. Consequentialist and character-based objections are considered, along with cases involving selfish and mixed motives. The paper concludes that doing the right thing for an operative reason the same community condemns is never wrong at the level of the deed but always wrong at the level of the will. This dual judgment also reveals a practical blind spot in shared morality: communities can condemn motives they often cannot observe or prove.

Keywords: Intentional Action, Moral Motivation, Collective Endorsement, Moral Relativism, Consequentialism, Moral Worth

Introduction

Consider a police officer who fires at a man advancing towards him with a knife. While this shooting is clearly self-defense, the officer also did it because he hates the man's race. Since both descriptions are accurate, the morality of the act becomes less clear. The act may appear completely permissible or vile and cruel. A common temptation is to treat these descriptions as opponents and see which wins. However, these two descriptions do not show a defect in moral perception. They give a clue to the underlying question. The answer to whether it is ever wrong to do the right thing for the wrong reason is *both never and always*. This apparent contradiction dissolves once the two perspectives are distinguished, and the question quietly blurs the line between them. This discussion will not attempt to settle whether morality is objective, that is, true apart from what anyone thinks. Instead, it uses the collective endorsement of a community because its settled standards can be applied to both the deed and the reason. Where the community endorses the deed and condemns the motive, the deed is never wrong, but the will is always wrong. A community that endorsed the officer's hatred would judge his motive differently.

Clarifying the Central Terms

The deed is what a person does, described apart from the motive. The will is the person's choice to act for a particular reason. "The right thing" refers to a deed that the community endorses when it is described without the motive: stopping a knife attack, saving a child, or returning a wallet. "The wrong reason" refers to an operative reason that the same community

condemns, not merely a reason that is less noble than another possible reason. For example, a person who gives to charity because the cause matters and because the donation brings satisfaction is not automatically acting wrongly. The problem begins when the reason expresses contempt, indifference, exploitation, or some other attitude the framework rejects. This matters because otherwise the argument would turn every imperfect motive into a moral violation. It would also make the answer trivial: a wrong reason would be wrong simply because the phrase calls it wrong. The claim is narrower and more important. Even when the outward deed is right and the operative reason is independently established as wrong, it must still be determined what the reason's wrongness attaches to: does it merely lower someone's opinion of the person, does it erase the deed's rightness, or does it make the same action answerable to two judgments? The rest of the argument defends the third view.

The Structure of Intentional Action

Everything depends on the nature of action itself. An action is not just an event in the world or some movement of the body that happens to come from a person. When someone acts, there is the deed and the reason behind it, and a common first instinct is to treat them as two separate things: the reason comes first, then the act follows. That misses the connection between them. The motive becomes the will, and once the will is carried out, it is the act. They are not two separate things, where the reason is one thing and the act it produces is another. People are used to separating the reason from the deed and judging each one on its own. But the reason is not attached to the action from the outside; it is part of what the action is. Take the reason away and you do not have the same action. You have a different action, or no real action at all, simply a body moving. They are one thing seen from two sides, first as something being decided, then as something being done. Once you see this, something follows. A single action can be described in more than one way, and all of those descriptions can be true at once, even though the action is only intentional under some of them (Anscombe, 1957). Consider again the case of the officer. His one act is truly "defending himself from a knife," and it is also truly "killing a man out of racial hatred." Both are accurate, and neither cancels the other. This is what makes the question possible. It does not ask whether the deed is wrong. It asks whether it is wrong to do the right thing for the wrong reason. That question names the act under two of its descriptions at once and asks to which one the wrongness belongs.

Anscombe helps explain why this is not just wordplay. To identify an intentional action, one can ask the agent "Why?" and trace the descriptions under which the question has an answer. A movement of a finger may also be pulling a trigger, stopping an attacker, protecting oneself, or taking an opportunity to kill someone one hates. The physical event remains one event, but not every true description has the same relation to the will. If the gunshot startles a bird, "startling a bird" is true of what the officer caused, but it is not necessarily something he did intentionally. By contrast, if he acts both to protect himself and to satisfy racial hatred, each reason helps explain why that shot occurred. Davidson's account of reasons as causes strengthens the point: an operative reason does not merely decorate the act after the fact; it helps produce and rationalize it (Davidson, 1963). This also sets an important limit. The officer may privately be racist without the racism being a reason for this shooting. If he would have responded identically to any attacker and the hatred played no role in his decision, then "shooting from racial hatred" would be a false description of the action even though it might be a true description of his character. The case works only because it stipulates that the hatred was among the reasons on which he acted. Once that is granted, removing the reason changes not merely an observer's attitude toward the officer but the intentional structure of the shooting itself.

Morality as Collective Endorsement

Where the wrongness lands depends on what "wrong" even means. On this account, morality is not objective on the scale of the universe. At that level, it is subjective, and nothing in reality makes an act right or wrong on its own. But that is not the end of it. A group of people can agree on a framework, and inside that framework you get a kind of bubble of objectivity. Right and wrong become real again. An act or a reason is right if the community would approve of it, and wrong if it would condemn it (Gowans, 2021). Think about a UFC fight. Nowhere in the universe is there a law against elbowing someone in the back of the head, and on a cosmic scale that move is not wrong, because nothing is. But if a fighter does it in an actual UFC match, a spectator who says he did something wrong is objectively correct. The sport agreed on its rules, and inside that agreement breaking them is a real fact, not just a spectator's opinion. The rules are made up, and the wrong is still real. Philosophers call a fact like this an institutional fact. It holds only because a group treats it as holding, but once they do, it holds completely (Searle, 1995). Morality is the same. All that is needed is a shared framework. So calling the officer's reason wrong is not just a matter of personal opinion. Inside the shared framework, it is a fact. There is a cost. In a community that endorsed the officer's hatred, his willing would not be wrong by that community's standard. It does not make the racism right in any way worth wanting. It means that a community committed to equal treatment condemns the racist framework from inside its own. Conventionalism does not make every framework equal. It only makes right and wrong framework-bound.

Collective endorsement must also be distinguished from a public-opinion poll. A temporary majority can applaud something without making it part of the community's considered moral framework, just as spectators can cheer a foul without changing the rules of a sport. What matters is the network of norms that members treat as standards: the reasons they offer one another, the expectations they enforce, the criticisms they recognize as intelligible, and the principles they appeal to when their own conduct is challenged. The word "would" in "would approve" therefore matters. It points to judgment under the community's settled practices of justification, not to whatever answer receives fifty-one percent in a moment of anger. Scanlon's idea that morality concerns principles people can justify to one another offers a useful way to understand this practice, even though this view remains more conventionalist than his (Scanlon, 1998). It also explains how internal moral criticism is possible. A society may officially permit discrimination while also claiming that citizens deserve equal concern. Reformers can expose the conflict by appealing to the framework's deeper commitments rather than importing a rule from outside every human practice. None of this creates morality at the scale of the universe. It only shows that social construction is not the same as arbitrariness. Money, offices, promises, and laws depend on collective recognition, yet particular claims within those practices can still be mistaken. Moral claims can work the same way. Their authority is made by a community, but once made, it can reach members who deny, misunderstand, or secretly violate it.

The Limits of Framework-Dependent Morality

A harder objection presses the claim that a racist framework is worse than one that rejects racism. If all wrongness is framework-bound, how can that comparison be more than preference? The answer should be honest: "worse" is also spoken from within a framework. A community can condemn a racist code by standards such as equal concern, reciprocity, protection from arbitrary harm, and consistency between the treatment one demands and the treatment one permits. Members of the other community may reject those standards, in which case no framework-independent court remains to force agreement. That is a genuine cost of the view, but not a contradiction. Practical condemnation does not become meaningless because it lacks a cosmic referee. Communities argue across frameworks by uncovering shared commitments, showing internal inconsistency, exposing harms, and persuading

people to revise what they endorse. Sometimes this succeeds because the frameworks overlap; sometimes conflict remains. The argument needs only the more modest point that, in the community relevant to the examples, racial hatred and indifference to a child's life are condemned reasons. Once that premise is fixed, the question about deed and will can be answered without pretending to solve metaethics in full. The account is therefore conditional but not empty: given a shared moral standard, it shows where its verdicts fall when one intentional act has both an endorsed and a condemned description.

Judging the Deed and the Will

Within this framework, the community is the judge, but it does not only judge what a person did. The community also judges why. People think worse of someone who helps a friend just to look good than of someone who helps because he actually cares, even when the help is exactly the same. So the community passes judgment on the reason behind an action, not only on how the action turns out. An action is the will and the deed together, one thing across two stages. You can look at just the deed, which is the part that reaches the world, or you can look at the whole action, including the reason behind it. For instance, a deed may be stopping a knife attack, which the community approves. He did the right thing. Now look at the whole action, reason and all. You see a man who killed because he hated another man's race. The community condemns that. He did the wrong thing. These two answers do not contradict each other. They are answers to different aspects of the same action. One looks only at the surface, where the deed is revealed. The other looks all the way down, to the reason it came from. Someone might object that the community cannot actually see the reason, so in real life only the deed can be judged. That does not hurt the argument. It is exactly what it predicts. The reason sits inside the person, where others may never see it, and only the deed comes out into the open. But a reason being hidden does not make it any less wrong. The community's standard condemns acting out of racial hatred, and it keeps condemning it whether or not anyone catches the officer's reason. The rule sets the wrong, not who happens to be watching. The wrongness is there the whole time. It is just out of sight. So the action gets two answers. As a deed it is never wrong and as a will it is always wrong.

The Consequentialist Objection

The strongest rival to this view comes from consequentialism. Mill argues that the motive does not determine the morality of an action, although it affects the worth of the agent; the person who rescues someone from drowning for payment still performs the right act (Mill 1861, chap. 2). That distinction initially seems to defeat this claim. It appears to give the clean answer that the deed is right, the person is morally unattractive, and nothing more needs to be said. Yet Mill's example also reveals why the issue depends on how much of the action is described. Wanting payment may be a further aim that leaves the rescue itself intact. Racial hatred in the officer case is different when it enters what the officer wills toward the person before him. He is not merely protecting himself while possessing an unrelated bad trait. He is also using the attack as an occasion to kill someone he values less because of race. Under that fuller description, the reason helps constitute the interpersonal meaning of the act. Ordinary moral judgment already recognizes this. Giving someone help in order to humiliate them, telling the truth in order to cause needless pain, and offering an apology only to manipulate an audience can produce outcomes that are useful or superficially correct. Still, the reason changes what one person is intentionally doing to another. The wrongness does not flow backward and make the protected life or prevented injury bad. It belongs to the action under the description that includes the agent's will. The consequentialist is therefore right about one level: the deed considered by its public effect remains right. The mistake is treating that level as exhaustive. A complete moral description must include not only what changed in the world, but also what the agent was trying to do through that change.

Wrong Reasons and Mixed Motives

The officer's reason is obviously evil, but the argument should hold for smaller reasons too. For example, consider a situation where a driver swerves to avoid a child in the road, but only does so because he does not want to dent his car. The reason is not evil, merely selfish. Is the willing still wrong here, or only in the ugly cases like the officer? It is still wrong, and the reason he gives is what reveals it. The example specifies that he swerved only to avoid damaging his paint. Take the case at its word: the child alone was not enough to move him. A will that places a child's life at nothing is not a neutral will; it is a bad one, even when it happens to produce the right deed. It convicts him the same way the officer's hatred convicts him, just more quietly. Kant made a milder version of this point. His honest shopkeeper, who deals fairly only because it pays, earns no moral credit (Kant, 1785/1998, 4:397). The claim here is stronger. The selfish reason does not just fail to earn credit; it shows a will that is actually bad. So a wrong reason does not have to be evil; plain selfishness is enough. A real driver usually has mixed reasons, and if he also cared about the child, his will is fine and the case is not about him. The critique only applies when the selfish reason is the one he truly acted on.

This is why "wrong reason" cannot mean simply "not acting from duty." Kant's shopkeeper lacks moral worth because prudence, rather than duty, explains his honesty, but a lack of special credit is not always the same as positive wrongness. Later work on moral worth makes the point from the other direction: a person can respond well to morally relevant features of a situation without consciously applying a moral rule (Arpaly, 2003). Someone who instinctively protects a frightened stranger out of compassion may act admirably even if the thought "this is my duty" never occurs. The argument does not require purity, self-denial, or a single abstract motive. It requires only that the reason actually driving the action be one the shared framework condemns. Self-interest is therefore not automatically disqualifying. A shopkeeper can care about profit while also recognizing that customers should not be deceived. A donor can enjoy praise while still being moved by the people the donation helps. The driver case is harsher because the stipulation says that the child's life had no motivational weight at all. If the car could not be dented, he would not swerve. That counterfactual helps identify the operative reason. Mixed motives should be treated in the same way. If concern for the child would independently have been enough to move the driver, the selfish consideration does not make the will wrong. If saving the paint is the only reason, the good outcome is a piece of luck attached to a will that fails the community's standard. The "always" in this answer therefore applies to genuinely wrong operative reasons, not to every human mixture of duty, affection, fear, pleasure, and self-interest.

Is the Wrongness in the Action or the Agent?

Another objection grants everything said so far. Someone could accept that an action includes its reason, accept the framework, and still say the argument cheats. They would say an action can only have one moral status, and the officer's action is right because he stopped a man with a knife. What the argument calls the wrong willing is not the action being wrong at all. It is just the officer being a bad person. So there is no real paradox. The act was right and the man is bad. On this objection, the argument has taken a bad person and dressed him up as a wrong act to invent a puzzle that was never there. This is the best version of the objection, and the answer is that it hides the wrong in the wrong place. The objector puts all the badness into the agent and calls him a bad man. But the officer may be kind to everyone except the man he shot. What the community condemns is not his character. It is this shooting, and the reason he did it. When people say "what you did was wrong," they are grading a deed, reason and all, not a soul. The two-target view has to rewrite that plain sentence into a claim about the person. That is unnecessary if the reason is part of the act. And the question proves the point. It does not ask whether the deed is wrong, but whether it is wrong to do the right thing

for the wrong reason. The reason is written into the thing being judged. So judging the reason does not slip in a separate verdict about the officer's character. The reason is the exact thing the question asks about. The two answers were in the question's own words from the start.

The objection also assumes that moral predicates must attach to an action in only one undifferentiated way. But one event can support several noncompeting evaluations because the descriptions select different features. A medical operation can be technically successful and nonconsensual; a promise can be legally valid and cruelly extracted. Saying one does not cancel the other because the predicates answer different questions. The same structure applies here. Let the narrow description be D: stopping a knife attack. Let the fuller intentional description be W: stopping the attack while acting from racial hatred. The community endorses D and condemns W. This does not mean that the same proposition is both true and false. It means that one action token falls under a permissible description and a condemnable description. The dramatic answer "never and always" is therefore not a logical contradiction. "Never" concerns the deed already stipulated to be right; "always" concerns the willing already stipulated to arise from a wrong reason. Nor does this reduce W to a report about character. Character is a relatively stable pattern across situations. The judgment here can remain local. An otherwise decent person may perform one act from spite, just as a generally selfish person may once act from real concern. What is condemned is the attitude expressed in this encounter. That is why an explanation can change what apology is owed. "I helped you only so that everyone would see your weakness" does not merely reveal an unpleasant personality. It reveals something additional that was done through the help.

The Practical Importance of Motives

There is also a practical reason to preserve both levels of judgment. Public institutions usually have better access to deeds than to motives, and they should not pretend otherwise. Evidence, due process, and the danger of mind-reading give law strong reasons to focus on observable conduct unless an intention is relevant and provable. Moral judgment, however, is not identical to legal punishment. Friends, victims, and agents reflecting on themselves often know more about the reasons behind an act. They may rightly feel gratitude for a benefit while also rejecting the contempt or manipulation through which it was delivered. The two judgments can coexist without requiring anyone to deny the good outcome. More importantly, motives affect the stability of conduct. A person who does the right thing only because cruelty happens to be profitable will stop when the incentives change. A person who recognizes why another life matters has a reason that travels across cases. A community concerned only with visible compliance may therefore produce correct behavior while leaving its norms fragile. Moral education cannot be limited to arranging rewards and punishments; it must also cultivate the kinds of reasons from which members act. None of this licenses punishment for unexpressed thoughts. It shows instead why the hidden part of action matters even when institutions cannot reach it. The deed determines what happened to others. The will shows whether the agent's practical relation to them conformed to the same moral framework.

Conclusion

All of this leads to one strange result, and it is not the one the question was looking for. The community is the author of right and wrong. It can approve of the officer's deed in stopping the attack and still condemn his reason for firing. That judgment of his will holds only in a community that rejects racial hatred. But the reason is the one thing the community usually cannot see. Now and then a reason surfaces. The driver who saved the child to spare his paint did something wrong in acting out of indifference, though someone who saw him swerve may never know why he did it. So the community sets the standard that condemns the will, and still may have no way to reach what it condemned. That is the blind spot every shared morality carries. When only the deed is known, the community has to leave its judgment of the driver's will open.

References

Anscombe, G. E. M. (1957). *Intention*. Basil Blackwell.

Arpaly, N. (2003). *Unprincipled Virtue: An Inquiry into Moral Agency*. Oxford University Press.

Davidson, D. (1963). Actions, Reasons, and Causes. *The Journal of Philosophy 60*(23), 685–700. https://doi.org/10.2307/2023177

Gowans, C. (2021). Moral Relativism. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2021 edition). https://plato.stanford.edu/entries/moral-relativism/

Kant, I. (1998). *Groundwork of the Metaphysics of Morals*. (M. Gregor, Trans. & Ed.). Cambridge University Press. (Original work published 1785)

Mill, J. S. (2002). *Utilitarianism* (G. Sher, Ed.; 2nd ed.). Hackett Publishing. (Original work published 1861)

Scanlon, T. M. (1998). *What We Owe to Each Other*. Belknap Press of Harvard University Press.

Searle, J. R. (1995). *The Construction of Social Reality*. Free Press.