

Justification and Bad Motives

Lawrence Crocker*

I. INTRODUCTION

The relation of a defendant asserting a justification defense to the purportedly justifying circumstances has come in for a good deal of scholarly attention. In particular, it has been widely debated what sort of warrant a justified defendant must have that the justifying circumstances exist. The most popular view is that, although she need not be correct, the defendant must have an objectively reasonable belief in each of the justifying circumstances: for example, a reasonable person with Alice's information would have concluded that the victim was a burglar coming through the window, even if it was really Alice's husband trying to unstick the sash.

A good deal less attention has gone to the question just what relation the justifying circumstances must have to the actor's actual motivation. Suppose Bea had loaded her pistol with intent to hunt down her enemy Chuck. On her way from the kitchen towards the front door, Bea saw Chuck coming through her dining room window. What if she fired, with the reasonable belief that Chuck was making a forced entry under circumstances giving her a privilege to use deadly force, but with ulterior intent to kill her enemy?

I believe that the answer that would usually be given to this question is wrong, and that the discussion of the question has not yet gotten very deep. It is easy to see why this question would receive little attention from the working bar. In Bea's case, the trial will almost certainly turn on the basis of her proffered belief that the man at the window was a burglar. If Bea, prudently, declines to speak with the police before retaining counsel, the world will likely never learn that her intentions had everything to do with the fact that she thought the man climbing over the sill was Chuck, and nothing to do with her belief that he was a burglar. Defendants who claim a justification always assert that what they were acting from was that justification, and justification statutes rarely separate out an "acting from" element.

George Fletcher contends, plausibly, if without much cited authority, that it is the "consensus" of Western legal systems that a justified defendant must act with

* Senior Lecturer, Dartmouth College Department of Philosophy; former Assistant District Attorney, New York County; defense counsel, Slotnick, Shapiro & Crocker. This essay was stimulated by a paper by my student Scott Glabe, *Necessary Conditions for Justification: Actor Knowledge and the Causal Link Between Belief and Action*, 2 DARTMOUTH C. UNDERGRADUATE J.L. 19 (2004). Many of my key arguments grew out of discussions with him, although we often reached different conclusions. I am also grateful for discussions with Julia Driver, Bernard Gert, Walter Sinnott-Armstrong, Roy Sorensen, Tay Stevenson and members of the Dartmouth Sapientia Colloquium, Spring Term 2005.

“justificatory intent.”¹ Even though it will affect few real world outcomes, it is worth seeing why the consensus is mistaken as a matter of the underlying morality of criminal justification.

II. THEORIES OF MOTIVES IN JUSTIFICATION

A. *Robinson’s Mugger*

Motivation has been actively debated in the case of an accused who has no inkling of justifying circumstances. Paul Robinson’s well-known and provocative view is that I am justified in punching a mugger with a club, even if my belief was that he was a jogger carrying a flashlight, and my motivation was anti-jogger.² This was a debate well worth having, and Robinson has held up his side of it with imagination, rigor, and tenacity. In the end, however, few of us were convinced.³ An unfortunate byproduct of the focus on “unknown justification” is that the “known but non-motivating” justification case has been all but lost in the shuffle. It is too quickly thought to go the same way as Robinson’s jogger. In fact, as I will argue, the known and unknown justification cases for bad-motive actors lead in two quite different directions. An examination of why they diverge reveals some important features of the logic of criminal justification.

It is worth reviewing why Robinson is usually thought to be wrong about the jogger case. We do not believe that a perpetrator has the moral high ground underlying the justification defense if he could not, at least in principle, set out at the time of his otherwise unlawful act, grounds that made it (at least apparently) acceptable. This principle was announced in *Queen v. Dadson*⁴ over a century

¹ GEORGE P. FLETCHER, *RETHINKING CRIMINAL LAW* 557 (1978). The Model Penal Code, which Fletcher cites in support of the consensus, actually runs both ways on the point. Section 3.02(1) gives the defendant a justification so long as he believes his conduct “to be necessary to avoid a harm or evil to himself or another.” MODEL PENAL CODE § 3.02(1) (1985). Clearly a defendant could so believe even if the conduct is the result of other intentions of his altogether. Proviso (a) to 3.02(1), however, requires that “the harm or evil sought to be avoided by such conduct is greater than that sought to be prevented by the law defining the offense charged.” This suggests that the actor must be acting with the purpose of preventing the evil, not merely the belief that he would prevent it. The self-defense section, 3.04, on its face requires only belief in the justifying circumstances, but § 3.05, defense of others, refers back to § 3.04 to give a justification for such force only if one could use “such force to protect himself.” (emphasis added).

² PAUL H. ROBINSON, *STRUCTURE AND FUNCTION IN CRIMINAL LAW* 102, 108–111 (1997). For Robinson, however, this unknowingly justified attack on the mugger might subject the defendant to liability for an impossible attempt. *Id.* at 111.

³ Larry Alexander, although disagreeing with Robinson’s position on unknown self-defense, would honor unknown justification in the case of lesser evils. Like Robinson, however, he would temper the unintuitiveness of this result by making attempt liability available in these cases. Larry Alexander, *Lesser Evils: A Closer Look at the Paradigmatic Justification*, 24 *LAW & PHIL.* 611, 630–34 (2005). Unknown consent is often thought to give a good defense. So if consent is thought of as a justification, Robinson can count on some support on this point.

⁴ 4 L.R.C.C.R. 35 (1850).

and a half ago, and has been the accepted judicial view ever since. In that case, Constable Dadson shot a fleeing wood thief unaware that the thief had two prior convictions for the same offense. Because the thief was a two-time loser, and hence a felon, it would have been lawful for Dadson to use deadly force in his apprehension had he known the facts. Because of his lack of knowledge of the justifying circumstance, Dadson's conviction for the death of the thief was upheld.

When we are assessing the blameworthiness of an action, as we surely are in adjudging criminal liability, it is the information in the possession of the actor at the moment of acting that controls. Yesterday's action may have turned out well. It may, in terms of consequences, have been the right thing to do, but considerations I can first give today do not reach back to lessen my blameworthiness yesterday. We do sometimes note that something "turned out all right" in moderating our criticism of a bad choice. That this typically only moderates the criticism shows that it is not a justification in the central sense of that term. If I say "he justified his action by showing that it turned out all right," I may well put scare quotes around "justified." The conceptual impossibility of unaware justification is not peculiar to legal and moral justification. My epistemological justification for believing something yesterday cannot be a function of information that I acquired only today. This far I think common sense is sound and Robinson wrong.⁵

The question remains, what further relation must the justifying circumstances have to the defendant for there to be justification? The answer, which may initially seem counterintuitive, is that nothing more is ever required. It is enough that the defendant believed that the circumstances obtained at the time of the otherwise unlawful act,⁶ and, in accordance with the law of most jurisdictions, that this justifying belief was objectively reasonable. It is not necessary that the justifying belief motivated him, that he acted from it, or even that it entered into his deliberations. It need not have been the reason or even one of the reasons for the defendant's act either that the circumstances existed, that they were morally or legally justifying, or that they would affect his chances of being arrested or convicted. One can have a good justification of self-defense even though one does not act *out of* self-defense. Bea's knowledge that a burglar was coming through the window should give her a complete defense in those jurisdictions that permit deadly force against burglars, no matter what role was played by her hatred of Chuck. Similarly, had he known that the fleeing suspect was a felon, Dadson should have been acquitted, whatever else was going on in his mind.

⁵ Insofar as they are explicitly or implicitly committed to unknowing justification, this error is shared by Alexander, *supra* note 3, at 630–34, and Judith Jarvis Thomson, *Self-Defense*, 20 PHIL. & PUBLIC AFFAIRS 283 (1991).

⁶ There is a minor exception in certain cases in which the defendant has manipulated the justifying circumstances for the purpose of providing a defense for a premeditated crime. *See infra* note 17 and accompanying text.

B. Fletcher's Weak and Strong Requirements

Robinson's best-known opponent on the unknown justification issue is George Fletcher, and Fletcher's arguments bear as well on known but non-motivating justificatory circumstances. Fletcher argued in *Rethinking Criminal Law* that the defendant must "merit special treatment"⁷ in order to have a justification. Therefore, "he or she must at least know of the circumstances supporting the claim"⁸ Fletcher then set out stronger and weaker versions of the connection between justifying circumstances and the justified action: "the stronger version would require that the actor be motivated exclusively by the justificatory criteria; the weaker version, that he merely be aware of them."⁹

In *The Grammar of Criminal Law*, Fletcher has filled in his argument in a way that unequivocally rules out the weaker version. There he endorses Aquinas's analysis of self-defense in terms of the doctrine of double effect. The intent of the self-defender must be to protect herself. The harm to the aggressor may be only an anticipated, not an aimed for, causal consequence.

Some philosophers write casually of a right to kill in self-defense cases. This idea is mistaken. There is no right to kill but, consistent with the principle of double effect, merely a right to fend off the attack: the death of the assailant is a permissible side effect. The basic principle is that a private individual may not intentionally kill another human being when the explicit object (rather than the side effect) of the action is to cause death.¹⁰

This initially attractive approach to self-defense will not generalize to all cases of justification. The public executioner and the soldier intend death as an explicit object, not a side effect. The doctrine of double effect would also forbid my use of necessary deadly force against two innocent aggressors, as the doctrine has it that even unintended side effects must not be worse than the intended good.

The crucial problem with applying the doctrine of double effect to justification, however, is brought out by the following hypothetical: Police officers investigating a drug house are fired upon from the other end of the hallway. Returning fire, and aiming, as per protocol, at the trunk of the assailant, the officers succeed in stopping the attack. Not unexpectedly, the assailant dies before medical

⁷ FLETCHER, *supra* note 1, at 565.

⁸ *Id.*

⁹ *Id.* at 559. A third version, in footnote, would require the actor "conscientiously" to weigh the competing considerations, at least in the case of the necessity justification. *Id.* at 559–60 n.21.

¹⁰ 1 GEORGE P. FLETCHER, *THE GRAMMAR OF CRIMINAL LAW: AMERICAN, COMPARATIVE, AND INTERNATIONAL* 14 (2007). For Fletcher, a good justification negates wrongdoing or wrongfulness. *Id.* at 51, 95. Putatively justifying circumstances that were no part of the intent or motivation for the criminal act could not do so.

backup reaches the scene. An autopsy reveals that shots from each of two officers were causally involved in the death.

Suppose that the assailant was unknown to Officer Smith. He was simply a presumed drug dealer firing at uniformed officers. The assailant was well known to Officer Jones, however. Suppose, now, one of three alternatives: (a) Jones hated the assailant and would be pleased if he were dead, but would not attack him unlawfully; (b) Jones had decided to kill the assailant if he had the faintest hope of legal justification for doing so; or (c) Jones had decided to kill the assailant when next he saw him. Suppose for each case that Jones fired with the described state of mind, though fully aware that his shot would have the effect of protecting himself and Officer Smith. For cases (b) and (c), this would be a side effect.

Clearly Smith is justified, and his actions fall squarely within the doctrine of double effect. Although the doctrine is sometimes stated in such strong terms that Jones (a) would not be covered, this is surely too demanding a reading of the doctrine. The pregnant woman who elects a hysterectomy solely to save her own life is not liable under the doctrine for the death of her fetus, even if her preference would be that the fetus not survive. Jones (b) pretty clearly falls on the wrong side of the doctrine of double effect as he is aiming at the death of the assailant, and Jones (c) is unequivocally condemned by the doctrine.

Although Jones (c) intentionally killed the assailant, it is an important feature of this case that his behavior would have been just the same if he had been a pure defender of self and other. In fact, if Jones had realized in the instant before he squeezed off his first shot that it was not his old enemy after all, squeeze he would have just the same. He was a bad motive killer but one with an automatic and inevitable fallback if the bad motive evaporated. Fletcher would have to admit that this fallback would be paradigmatic self-defense.

In the real world, of course, the question of prosecuting Jones is very unlikely to arise. Let us suppose, however, that Jones, in a fit of self-awareness and guilt, confesses to the police review panel that the circumstances were those of (c). Ought Smith receive a commendation and Jones a murder indictment?

Although he displayed bravery in entering the drug house and did help save the life of a fellow officer, Jones's deplorable motive is enough to rule out a commendation. This much we can cheerfully concede to the doctrine of double effect, the point of which is that aiming at evil is morally worse than aiming at good while recognizing the evil as an inevitable side effect. Does the doctrine have the additional force, as Fletcher would have it, of taking Jones outside the moral justification of self-defense? Was what he did morally wrong, though his conduct was physically identical to Smith's and his mental state would have been consistent with paradigmatic self-defense and defense of others if we subtract his specific animus against the assailant? More seriously, does the doctrine of double effect mandate denying Jones (c) the criminal defenses and subjecting him to a conviction for murder? Neither Fletcher's discussion, nor, so far as I can see, the traditional resources of the doctrine of double effect, are sufficient to establish that point.

C. Dillof's "Regarded" Reason

There is obviously a good deal of conceptual room between Fletcher's strong "exclusive motivation" condition and his weak "awareness" condition. The stronger condition is so strong that it would convict Jones (a), and make the public executioner liable for murder if he happened to hate the person to be executed, or even if he took pleasure in his work.¹¹ The weak condition, requiring only awareness, seems very defendant-friendly, and has been widely rejected. Larry Alexander is one of the few exceptions.¹² I will argue, however, that, not only is the strong requirement too strong, we should not compromise on any of the possible intermediate positions between Fletcher's strong and weak conditions.

Anthony Dillof proposed such an intermediate position in his detailed and systematic examination of the relation of reasons, intent, and motivation to criminal justification in a 2002 article. Dillof defined his position as follows:

Let us say . . . that an actor "regards" the justificatory circumstances surrounding his conduct where (1) he believes justifying circumstances exist, and (2) based on his belief, he decides to engage in the conduct. "Regard" may be thought of as shorthand for the more bulky phrase "takes into account in his reasoning process when deciding what to do" or "functions as a reason (even if not the sole reason) relevant to the actor's choice."¹³

There are some ambiguities lurking here. Parts of the formulation suggest that merely considering the circumstances or according them positive weight might count as "regarding" them.¹⁴ But Dillof's examples make it fairly clear that a "regarded" reason is to be a "but for" reason for the decision and a "but for" cause of the action. The action is to be "based on" the justifying circumstances in that knowledge of those circumstances must have "made a difference."¹⁵

¹¹ The executioner example has probably been current in discussions of justification for many years. I made use of it in earlier drafts of this paper from 2004. Larry Alexander used it to this same end. Alexander, *supra* note 3, at 620. See also Vera Bergelson, *Rights, Wrongs, and Comparative Justifications*, 28 CARDOZO L. REV. 2481, 2486 (2007). Bergelson would restrict the point to the "public authority" family of defenses. For other sorts of justification she would require the perpetrator to have "justificatory knowledge and purpose." *Id.* (emphasis in original).

¹² Alexander, *supra* note 3, at 619. Alexander never requires motivation, and requires awareness in self-defense cases, but not in lesser-evil cases.

¹³ Anthony M. Dillof, *Unraveling Unknowing Justification*, 77 NOTRE DAME L. REV. 1547, 1597-98 (2002) (footnote omitted).

¹⁴ *Id.* at 1598. Dillof's formulation raises over-determination questions. If the actor deliberates with respect to three reasons, any one of which would be sufficient for him to perform the action, but only one of which is justifying, it is not entirely clear whether the justifying action is "regarded." The examples, however, suggest it is not.

¹⁵ *Id.*

Dillof gives the example of Mike, whose longstanding desire to slug Vincent has been restrained by moral scruples. Upon seeing that Vincent is about to attack Mary, whom Mike would otherwise do nothing to aid, Mike rushes in, punches Vincent, and saves Mary. The last fact is quite irrelevant to his motivation, but essential to his decision. Mike would not have punched Vincent had Mike not known it would be justified as a defense of Mary. So, a “regarded” reason need not be a motivating reason, a primary reason, a substantial reason, or even an “affirmative” reason, in the sense that it would militate in favor of the decision in the absence of other reasons. Here, had it been anyone other than Vincent, the less than public-spirited Mike would not, in the least, have been tempted to intervene. The justifying circumstances served only to neutralize his moral scruples against punching Vincent.

III. JUSTIFICATION DESPITE BAD MOTIVE

A. *A Case of Bad Motive Justification*

To explore the relation of the justifying reason to the action justified, I am going to borrow Mike and Vincent, and put them through a variation of Dillof’s case. Suppose that Mike’s long simmering hatred of Vincent has finally reached the boiling point. Moral scruples be damned, he is going to get Vincent. Approaching Vincent, he notices that Vincent is about to attack Mary. Mike realizes that his knowledge of this circumstance brings him at least very close to a justification for punching Vincent. Deliberating quickly, and with perhaps a little more than typical sophistication, Mike concludes that he already has a justification on Fletcher’s weaker condition. Fletcher’s stronger standard, requiring exclusive motivation, is beyond his reach. He cannot shed his hatred-induced desire to punch Vincent.

What, if anything, can he do to bring himself within the safe harbor of Dillof’s “regard” analysis? He need not rework his motivation, as he would for Fletcher’s strong analysis. A regarded reason need not be a motivating reason. But Mike stumbles on the fact that his decision was made before he caught sight of Vincent and Mary. So the justifying circumstances were not “but for” reasons for that decision. Perhaps the situation can be saved, however. He thinks:

Well, then, let me withdraw that old decision and consider anew. The old decision was razor thin. My hatred barely overcame my moral scruples, and I knew what I was going to do was wrong. Now I can punch Vincent with a clear conscience. I have a reason that neutralizes my scruples. Instead of being a close decision, it is easy.

Being a careful reader of Dillof, however, Mike realizes that all this is insufficient for the “regard” analysis. For *that* analysis, he must conclude that he would not arrive at his new decision to attack Vincent in the absence of the

justifying reason. It is not, he thinks, enough that the justifying reason is necessary for the new decision described as an easy decision about which he has no moral doubts. In fidelity to the regard analysis, the justifying reason must be necessary to the minimally described decision to punch Vincent. If he would, for the old reasons, slug Vincent in any event, he does not “regard” the justifying circumstances.

Mike concludes that it all depends upon whether, in his new circumstances, he would still, counterfactually, decide to punch Vincent, even if Mary turned out not to be in peril. Somewhat frustrated, he wonders whether he can determine the truth of that counterfactual. It is tricky because Mike must take his present circumstance, and then mentally subtract the most significant aspect of that circumstance—that Mary is in fact in peril.

It might be suggested that Mike could unravel the counterfactual simply by noting that nothing had happened to lessen his hatred, which had been sufficient for him to punch Vincent before he learned of Mary’s plight. Yet if the hatred has not decreased, may the moral scruples have increased? If Mike’s earlier hatred-colored rationalizing had weakened his moral scruples, might they be brought back towards their normal level by the realization that Vincent could receive his comeuppance through the criminal justice system or through the use of justified force against him by Mike or another onlooker? Mike’s rejuvenated scruples might then survive his imagined subtraction of Mary from the situation, and he might conclude that now he would not slug Vincent if he were not also saving Mary.

Although this reasoning provides a conceivable way for Dillof’s analysis to lay claim to the right answer in this Mike-and-Vincent variation, it also shows that the reasoning is unappealing. Mike should neither have to evaluate this artificial and difficult counterfactual, nor “change his mind” in this way. The punch he is about to throw, with a newly clear conscience, is one that society will approve, for reasons he fully understands.

Insofar as one can endorse the views of hypothetical characters, I endorse Mike’s. Even Dillof’s analysis demands too much. The justifying information need not have been a reason for the defendant’s act.¹⁶

At trial the Assistant District Attorney’s cross examination of Mike might run as follows:

Q. You made up your mind to punch Vincent before you ever saw Mary, isn’t that right?

A. Yes.

Q. And Mary had nothing whatsoever to do with that decision, correct?

¹⁶ Many philosophical accounts of the relation of reasons to actions require a causal connection between the reason and the action. See Donald Davidson’s classic, *Actions, Reasons, and Causes*, 60 J. OF PHIL. 685 (1963). A consideration you could give in justifying your action need not be a reason for your action.

A. Yes.

Q. The sole reason you decided to punch Vincent was your hatred of him. Isn't that also right?

A. Yes.

Q. And you never changed that decision in any way, did you?

A. No.

Q. So the reason you punched Vincent, and the only reason, was because you hated him?

A. Yes.

Then follows the proverbial one question too many:

Q. And even when you came face to face with Vincent, you never wavered in your determination to assault another human being, did you?

A. Well, I didn't reconsider my decision. But then, I didn't have any reason to. I knew that Vincent was about to attack Mary and that by punching Vincent I would save Mary. I didn't really think of it as an assault anymore, because, it was, like, you know, justified.

Q. Move to strike as unresponsive, Your Honor.

COURT: Denied.

Defense counsel will use Mike's last response to good effect at the charging conference to persuade the judge to instruct the jury that Mike was justified if he was aware that Vincent was about to attack Mary and that punching Vincent was the least force that would prevent that attack. Counsel would be satisfied with this instruction, although she would press in addition for an explicit explanation that motive is irrelevant.

What I will next do is survey some of the arguments that Mike's lawyer might give at the charging conference. Like a real world defense counsel, I will not be punctilious about the underlying theories¹⁷ or their *bona fides*, until I reach what I take to be the crucial argument.

B. *Incentives*

Mike's attorney might well begin with incentive effects. This consequentialist argument for knowing but unmotivated justification is the one argument that has been developed in some detail in the literature through the efforts of Larry Alexander, in his discussion of the famous "Trolley Problem":¹⁸

¹⁷ To remedy this, see JOSHUA DRESSLER, UNDERSTANDING CRIMINAL LAW 222–25 (4th ed. 2006).

¹⁸ Philippa Foot first set out the "Trolley Problem" in *The Problem of Abortion and the Doctrine of the Double Effect*, in VIRTUES AND VICES AND OTHER ESSAYS IN MORAL PHILOSOPHY 23 (1978). The case involved the issue of the moral permissibility of a trolley driver's diverting his runaway trolley onto a spur track, thereby killing one worker on that track, because the trolley would

Suppose in Trolley—the original Trolley, with one siding—defendant has tentatively decided against switching the Trolley and saving the five, but then he recognizes the lone worker on the siding as someone he has always despised, changes his mind, and switches the trolley to the siding. He is no doubt a bad person. But he has done what is socially preferable. To punish him for his motivation will deter badly motivated people from socially desirable conduct whenever they would not engage in such conduct but for their misanthropic motivations.¹⁹

Similarly, Mike is not as virtuous a person as is Pat, who would have gone to Mary's defense for the sake of protecting an innocent victim. If Mike is a strapping longshoreman and Pat a puny lepidopterist, is that just too bad for Mary? Surely not. Both Mike and Pat recognize Mary's plight. Society would want Mike as well as Pat to have an incentive to come to use measured force against Vincent in these circumstances. So we have, as Alexander argues, policy reasons for granting the privilege to Mike as well as to Pat.²⁰

Consider a case of justification outside of the self-defense family. A stalker lays out a picnic lunch for himself and his intended victim on a secluded point past which the object of his attentions swims each day at noon. Despite an order of protection prohibiting him from any contact with her, and the risks of a kidnapping charge, his plan is to pull her into his small boat. He is sure that, by dint of his charm and her change of heart, they will have a wonderful time picnicking together. As he rows, he notices that she is having trouble swimming, and is about to go under. He continues with his plan. He is gratified when she thanks him for saving her life, but disappointed that she rejects his picnic, and even more disappointed when he is arrested. The charges should, however, be dismissed, even if he were so incautiously honest as to admit both his original plan and that it had remained unchanged in the new circumstances.

Although no kidnapping charge could be brought, because the "against her will" element²¹ was lacking as things turned out, there was a facial violation of the

otherwise kill five workers on the main track. Foot contrasted the intuition that the number of lives saved might make it permissible to switch tracks with the intuitively impermissible conscription of one involuntary organ donor to save the lives of five organ transplant patients. Judith Jarvis Thomson extended Foot's search for a moral theory to account for the intuitive difference between the cases, and in so doing changed the trolley decision maker from the driver to a bystander at the switch, thereby eliminating duties a driver might have as an agent of the trolley company. Judith Jarvis Thomson, *The Trolley Problem*, 94 YALE L.J. 1395 (1985).

¹⁹ Alexander, *supra* note 3, at 619–20.

²⁰ The incentive argument depends upon the availability of some reasonably simple principle encapsulating the privilege for bad motive actors. "Motives don't matter" would do. We would not get an incentive effect if the general population were required to use such precedents as Alexander's Trolley version as a guide to conduct.

²¹ Whether expressly or as interpreted from such statutory words as "restrain," kidnapping always has a "without consent" or "involuntary" element. *See, e.g.*, N.Y. PENAL LAW § 135.00 (2004); CONN. GEN. STAT. § 53a-91(1) (2007); *Chatwin v. United States*, 326 U.S. 455, 464 (1946).

order of protection. On these facts, however, that violation was fully justified. The defendant saved a woman he knew to be drowning. That is justification even if his consistent intent was to pull her into his boat for his own purposes.²² It is not in our interest as a society to deny the availability of the defense to those who were not intentionally, but merely knowingly, bringing about these better states.²³

It is not only that we want to encourage bad people to do things of which society would approve; we also want no disincentive for people of mixed and equivocal motivation, or people who are simply squeamish at the thought of having their motivations deeply probed by the police or prosecutors.

C. Deterrence

Alexander's incentive argument can be extended to the aggressor's side of the encounter. We need not buy too far into the "deterrence" theory of the defenses of self-defense, defense of others, and defense of property to conclude that, other things being equal, it is better that an unlawful aggressor fear that he will be the target of defensive counter-measures. If the objective circumstances support such counter-measures, there is some loss of deterrence if we deny legal immunity to improperly motivated potential defenders. There is nothing to be gained in lowering the deterrent threat by disallowing force against aggression by knights of less than the purest heart.

A cousin of the deterrence theory is the retributive theory of self-defense, defense of others, and defense of property. This theory is more popular among the general population than among theorists in English-speaking countries. However, if action against aggressors did receive some part of its justification from a retributive rationale, that would also support granting Mike the defense. As far as "natural punishment" goes, Mike is as good an instrument of Vincent's punishment as is Pat.

²² This hypothetical suggests the possibility that a defendant who had a good justification defense on the current analysis might yet be convicted of an attempt, the cousin of Robinson's projected attempt liability for the "unknowingly justified." Our stalker, for example, was arguably "lying in wait," usually considered enough of a substantial step to constitute an attempt. *See, e.g.*, MODEL PENAL CODE § 5.01(2)(a) (1985). An attempt conviction for preparations prior to an act that is itself legally justified is certainly possible. I regard this as a problem, but it is a problem in the prevailing law of attempts, not in the present theory of justification. It is a minor scandal that the stalker could, in principle, be convicted of attempted kidnapping when he carried all the way through a plan that did not constitute kidnap, because it was not "against her will." The law of attempts is filled with similar paradoxes created by overly inclusive definitions of attempts. *See generally* Lawrence Crocker, *Justice in Criminal Liability: Decriminalizing Harmless Attempts*, 53 OHIO ST. L. J. 1057 (1992).

²³ *See* DRESSLER, *supra* note 17, at 223.

D. *Forfeiture of Rights*

Mike's lawyer would also be sure to mention theories, like that of Judith Jarvis Thomson, on which the aggressor loses his rights. The "Irrelevance-of-Intention-to-Permissibility Thesis" that Thomson holds is nearly inevitable on such a theory: "It is irrelevant to the question whether X may do alpha what intention X would do alpha with if he or she did it."²⁴ At least in Mike's case, it is common sense that Vincent lacked a right against a measured use of force by anyone who knew that Vincent was about to attack Mary. The right was gone quite independently of Mike's motive. It is in this respect, like my privilege to ignore the instructions of Utah police, if I know that I am on the Arizona side of the border.

IV. THE IMPORTANCE OF THE POSSIBILITY OF RECONSIDERATION

Finally, defense counsel would turn to what seems to me the weightiest argument against denying Mike a justification defense. It may not be the easiest argument to bring into contact with the judge's existing categories, because it is not, so far as I know, recognized with any clarity either in decisional law or the literature of criminal jurisprudence. Mike's lawyer should argue, however, that it is simply unjust to hold Mike criminally liable when he knew that all the justifying conditions for his action existed at a time when he could have reconsidered and decided not to punch Vincent.

The prosecutor will object, of course, that any act done from an exclusively bad motive is a bad act, and the actor is a bad person even if information was available that could have supported a motive that would have converted it into a good act. She will assert: "To be morally relevant a potential reason must have at least some link to the motive and the action. It cannot simply be floating in the air."

I admit that bad motives are bad and that a person acting from a bad motive is, in that respect, a bad person. We all prefer Smith to Jones, Pat to Mike, and a pure-hearted lifeguard to the stalker. You will want to concede, however, that the actions of Jones, Mike, and the stalker were not unequivocally bad acts. They saved Smith and Jones, Mary, and the swimmer.

There are two points at which we may part company: (1) Can the bad motive actors take credit for the respect in which their acts were good?; (2) Does the respect in which their acts were bad show the actors to be dangerous people? For purposes of the criminal law, I argue that the answer to (1) must be "yes," and the answer to (2) "no."

²⁴ Thomson, *supra* note 5, at 294.

A. *Giving the Viciously Motivated Credit for Good Acts*

If someone commits what would otherwise be a crime in a mental condition in which she absolutely could not have reconsidered and changed her mind, she will not have performed a criminal act or she will have an excuse. She was pushed or blown onto the victim, or had a seizure, or was sleepwalking, or was insane. Being unable either to consider or to reconsider, she is not responsible for the harm, and will have no criminal liability.

If someone commits an act in circumstances that justify what would otherwise be a crime, the fact that he could have changed his mind is, symmetrically, something for which he is responsible and properly bears upon his criminal liability. Mike could have eschewed saving Mary, though he knew her to be in peril. He is morally responsible for not having changed his prior plans to Mary's detriment. The stalker is responsible for not deciding at the last moment to honor the letter of the protective order and to let the swimmer drown. Jones is responsible for not deciding against firing at his enemy, and thereby is responsible for saving Smith's life. In each case the actor was aware of the justifying conditions during a period in which reconsideration could have taken place. This should be enough, as a moral matter, to give him some credit for the good results of his actions.

It should not be surprising that the possibility of last-minute reconsideration is important in getting the right characterization of justification in the criminal law. The possibility of last minute changes of mind is morally relevant to praise and blame in various settings, even when no last minute reconsideration in fact takes place.²⁵

Consider: Deidre hears that her son is drowning in the heavy surf. She determines to attempt the rescue. Racing forward, and just as the first breaker crashes at her feet, she realizes that the child is not her son. She continues, undeterred by the ferocity of the surf, making the rescue. Deidre is praised for her selfless heroism in saving a stranger. That she did not reconsider when she saw it was not her son is clearly to her moral credit. Suppose someone objected that she should only get credit for a stranger-rescue rather than a son-rescue if she actually went through a process of reconsideration when she realized it was not her son. It is not simply her knowledge that must change; it must be a fresh decision with a new motive. Wouldn't we quickly silence this carping by observing that it was enough that she knew it was not her son when she could have stopped her charge into the surf?

The moral significance of not changing one's mind goes both ways. Had Deidre thought initially she was saving a stranger, and realized upon reaching the surf that it was her son, her action would still be heroic, but might be a little less praised because she knew it was her son when she could still have decided that the rescue was too risky to attempt.

²⁵ Crocker, *supra* note 22, at 1079.

Or, consider: Evan had long determined to sell one of his downtown buildings as soon as he got a decent bid for it. A community group gave him a barely acceptable bid to put in a methadone clinic. He went ahead with his plan even though his real estate consultants told him that the clinic might negatively affect the value of two of his other buildings. Evan should get credit for not reconsidering and changing his mind. In the reverse case, Evan goes ahead and sells to a warehouse developer, even though his rental agent tells him that his building currently houses a successful methadone clinic. For not changing his mind in these circumstances, Evan may properly be blamed.

The prosecutor might interject that in the cases of Deidre and Evan, the motive automatically and inevitably changes when they gain their additional knowledge that it is a stranger or that the sale will negatively affect other buildings. There need be no reconsideration as such for us to redescribe the motive to include the new information. Deidre intended to save her son, and her intent changed with her new awareness to an intent to save a stranger. I would agree with the prosecutor that hers is a possible description, but so is that Deidre intended to save the drowning person, only believing him to be her son. Her intent, so described, remained unchanged, even as her beliefs changed.

I see nothing wrong with redescribing Deidre's and Evan's motives in terms of their new information, but the same redescription is equally available to Jones, Mike, and the stalker. Jones intended to fire at an enemy whom he knew was trying to murder Smith. If this redescription is deemed to turn Jones's motive into a good one (at least sufficiently good for the criminal law defense), then my thesis will need reformulating in terms of "initially bad motives" or "mixed motives." Although I am unpersuaded that the sort of information in question requires, rather than permits, the redescription of the motive, my primary concern is that Jones, Mike, and the stalker get credit for not changing their minds, and so have the criminal defense, not whether they are described as having bad motives or modified bad motives.

It might be suggested that not changing one's mind upon gaining new information is something for which one can get credit in ordinary situations of moral praise and blame, but not such credit as to bring a defendant within a criminal law justification. This contention, however, is very much in need of an argument, an argument made difficult by the seriousness of denying a defense to a criminal defendant. If credit is, as a moral matter, normally available to those who do good acts with knowledge of the circumstances that make them good, it is, *prima facie*, all the more pressing that Officer Jones, Mike, or the lifesaving stalker be able to use that credit against the life-wrecking effects of a criminal conviction.

The contention that the credit they should get for anticipated good results is too insubstantial to affect the criminal law is also belied by the law's use of the symmetrical consideration to establish liability on the basis of new information within the ambit of possible reconsideration. Suppose that Ahab has been plagued by a large white pigeon that fouls the window ledge of his apartment, coos loudly at all hours, and mocks his cat. Ahab determines to shoot the pigeon. He misses

when the pigeon is walking on a roof and again when it is perched on an old television antenna. The pigeon then flies down to a park filled with people on lunch break. If Ahab fires at the pigeon now, he may well be guilty of reckless endangerment, or, if his shot goes astray, reckless assault, reckless manslaughter, or, conceivably, depraved indifference murder.

It will do Ahab absolutely no good to argue that he had no misanthropic purpose, his sole motive being to rid the world of the white pigeon. His circumstances changed completely as soon as he became aware that the pigeon was in a crowd of people. It is of no benefit to his defense that his motive and intent did not change. Indeed, it is because he did not reconsider and change his mind about shooting the pigeon when it flew down to the park that the recklessness element of the offenses is made out. Good motives are not enough to prevent criminal liability if the actor had new information that should have led him to reconsider, change his mind, and avoid causing risk. Why should the defendant not get the benefit in the symmetrical case of a bad motive with awareness of good consequences?

I want to emphasize that in arguing the moral significance of the possibility of reconsideration for the defense of justification, I am not retreating from my claim that motives should be entirely irrelevant when there is awareness of justifying circumstances. I am not suggesting that actor need actually reconsider, or even consider whether to reconsider. Just as all that is required for Ahab to be liable is that he could have reconsidered when the circumstances turned negative, the possibility of reconsideration when the circumstances become potentially justifying is all that is necessary to cut off criminal liability.

B. Dangerousness

Suppose that the judge now turns to the prosecutor and asks for argument. The prosecutor would be sure to contend that it is a purpose, perhaps the chief purpose, of the criminal law to protect the citizenry. One way the law protects is by incapacitating those who have, through their conduct, demonstrated their dangerousness.

Jones was motivated by hatred of the drug dealer, Mike by hatred of Vincent, and the stalker by passions unhemmed by sensitivity to the private space of his intended victim. Won't we be safer if they are all behind bars? Perhaps we will be. They may well be dangerous people. We do not, however, have a substantiation of their dangerousness of a sort that should be required by the criminal law. The problem goes much deeper than ordinary insufficiency of the evidence. It comes down to a counterfactual that cannot be resolved.

We would know that Jones was dangerous if he shot the drug dealer in the back while the dealer was eating a sandwich with friends. Mike's dangerousness would be sufficiently shown if he found Vincent alone in a back alley and pummeled him. We would have no doubts about the stalker's dangerousness if he hauled a strongly swimming victim into his boat over her vociferous protests.

That is not, however, how things were on our suppositions. We have no basis for confidence that the closest possible world in which the drug dealer does not fire upon Smith and Jones is one in which Jones nonetheless shoots him. In the world in which Jones actually fired, his life and Smith's hung in the balance. We do not have the luxury of running Jones through possible worlds to see in which he does, and in which he does not, reconsider his plans. Although in support of a very different argument, Alexander makes the following observation about counterfactuals involving culpable human action: "[T]he truth of such counterfactuals, if counterfactuals can be true, is too speculative a notion on which to base criminal liability."²⁶

This is a reprise of the point made in considering Mike's difficulty in evaluating the counterfactual he would be faced with on Dillof's compromise analysis. Neither Mike nor a jury can be confident what Mike would have thought or how he would have behaved had he not been aware that Vincent was about to attack Mary. We may indulge our speculations in deciding whether to befriend Mike, Jones, or the stalker. Indeed, assuming we know their motives, we would have enough information about them to be wary even if they had never encountered Vincent, entered the drug house, or gone to the shore, respectively. Their thoughts may be enough for our personal purposes, but not those of the criminal law.

C. Motive and the Necessity Requirement

The prosecutor might also argue that the necessity requirement, present in all defense-of-others statutes in one form or another, entails that the justification requires a proper motive. Mike may not use more force against Vincent than is reasonably required to save Mary. Therefore Mike must calibrate his use of force to Mary's circumstances. That is, these justifying circumstances, at least, must enter into Mike's deliberations.

The prosecutor, however, is wrong in arguing that the statute's necessity element entails that a justified actor must actively consider necessity. It is true that an actor who uses excessive force will only have the defense if she reasonably believed that circumstances were such that her force was not excessive. An actor, whether of good or bad motive, who uses force that is not excessive, however, need not have given it any thought at all. Some statutes facially require a belief or reasonable belief that the level of force is necessary under the circumstances, but in reality the defendant's state of mind will not be examined, and should not be examined, if her level of force was in fact necessary.

²⁶ Alexander, *supra* note 3, at 635. Alexander did not unequivocally endorse this skepticism about such counterfactuals, expressing it as part of an argument he ultimately rejected. Yet, he is sympathetic to this step for counterfactuals involving culpable human choice. In the rejected argument, the "too speculative" point is used against the defendant. In light of the burden of proof beyond a reasonable doubt, the use of speculative counterfactuals to establish criminal liability is far more suspect.

D. *Causing the Conditions of One's Own Defense*

A sufficiently scholarly prosecutor might argue that granting Mike the justification would require the acquittal even of defendants who construct situations in which there will be circumstances justifying their premeditated attacks on their enemies. This horrible scenario, however, is not entailed by my thesis. I am certainly not claiming that all bad-motive defendants who claim a criminal justification are entitled to the defense. Manipulating the justifying circumstances with bad motive will unquestionably sometimes be enough to take the actor out of the defense. This is the traditional problem of "causing the conditions of one's own defense."²⁷ I think we have at least a good general idea of what sorts of manipulations of justifying circumstances are incompatible with a good justification defense. Insofar as there is a residual problem, it is one equally for any plausible theory of the relation of motive to justification.

Consider Fletcher's restrictive "exclusive motivation" requirement. If we interpret this so that any bad motive anywhere in the act's history disqualifies the defense, then the theory will, properly, deny the defense of self-defense if the victim was manipulated into an attack. Unless special provision is made, however, the exclusive-motivation requirement would also deny the defense where it should be available. Suppose Oscar is determined to kill Elmer, puts a knife in his pocket, and heads out his door in Queens to hunt for Elmer in Manhattan. No sooner is he out the door than shots whistle by Oscar's head issuing from the barrel of Elmer's pistol, just feet away. As the shots continue, Oscar desperately launches himself at Elmer, and succeeds in plunging his knife into Elmer's heart. In order for Fletcher to absolve Oscar on the exclusive motive test, he would have to restrict the test to Oscar's immediate motive, excluding from consideration his ulterior bad motive.

If we so interpret the "exclusive motivation" test, however, its associated theory of justification will also need a special exception for manipulated justifying circumstances. The capable creator of the conditions of his own defense may genuinely have the immediate motive to defend himself. It may even be his sole motive at the moment of the fatal thrust. Specifying what sorts of prior manipulations of the circumstances are incompatible with the justification defense is an added problem for any plausible justification theory, mine no more than those with very different views of motives.

E. *Bad Motives and Lesser But Not Least Evils*

Alexander maintains that the justification defense should be available to a defendant who elects a lesser evil, but because of bad motive, not the least possible evil. He tenders the following variation of the Trolley problem.

²⁷ See Paul H. Robinson, *Causing the Conditions of One's Own Defense: A Study in the Limits of Theory in Criminal Law Doctrine*, 71 VA. L. REV. 1 (1985).

[T]here are two sidings on to which the [t]rolley can be diverted. On one, there is the sleeping lone worker. On the other, there are two workers, also asleep. And . . . on the main track there are five sleeping workers. Suppose further that the lone worker is a good friend of the defendant, whereas the two workers on the second siding are strangers to defendant, or even people he knows but dislikes. Suppose, finally, that defendant switches the trolley to the siding with the two workers rather than to the siding with the one worker.²⁸

Although conceding that the “defendant did not bring about the least evil possible”²⁹ and that the two deaths were neither necessary nor believed necessary to save the five, Alexander argues that if we tell the defendant that “he will not get the defense unless he kills the one, he will likely do nothing, and the five will be killed.”³⁰ He concludes: “The wiser policy thus appears to have the defense be available to one who chooses a lesser evil even if he does not choose the least evil of which he is aware.”³¹ This is the conclusion to which a strict incentives approach leads. A focus on the possibility of reconsideration will lead to a different conclusion.

Mitchell Berman crafted a telling variation of Alexander’s case:

[T]here is only one spur off the main track. On that spur, 500 yards ahead, bystander can see one sleeping worker. If bystander switches the trolley, he will save five workers at the cost of killing one. Unfortunately, that sleeping worker is bystander’s good friend. Fortunately, bystander can see that, 200 yards down that spur . . . there is a second spur. Unfortunately, bystander can see two workers sleeping on that spur. . . . [B]ystander does switch the trolley onto the first spur, thereby saving the five workers on the main track, but also thereby imperiling his friend. Before the trolley has traveled 200 yards, however, he switches the trolley onto the second spur, thereby saving his friend by killing the two sleeping workers.³²

Berman is right that in a prosecution for killing the two, this bystander would and should be denied the lesser evil defense, even if this means that there will be no incentive to throw the first switch to save the five. On my view this is because the possibility of reconsideration reaches right up to the no-return point of each decision in a chain of decisions. It inevitably makes each stage independently

²⁸ Alexander, *supra* note 3, at 618–19.

²⁹ *Id.* at 619.

³⁰ *Id.*

³¹ *Id.*

³² Mitchell N. Berman, *Lesser Evils and Justification: A Less Close Look*, 24 LAW & PHIL. 681, 687–88 (2005).

relevant to the justification. If the bystander's initial decision was to switch to save the five and then to switch to save his friend, he is responsible for not changing his mind when he came to the second switching.

Lesser evils are clearly not always good enough. Suppose that throwing the switch will save one hundred on the main track, throwing left will take the trolley onto a spur on which there are ninety-nine strangers, throwing right onto a spur with one stranger. Bystander decides to throw the switch left, because it is a little easier for a right-hander. He has saved one life over the baseline inaction, but cost ninety-eight lives more than need be. The evil was lesser, but so little so in comparison to what was possible that the defendant does not deserve benefit of the defense.

Yet, it seems doubtful that the best possible action should be required for the defense in every case. Suppose the bystander can save five by throwing the switch either way; left will jostle the passengers, and right will scrape the side of the trolley. If, being a trolley buff, he throws the switch left, he should not be denied the defense even if he knew that society would regard minor passenger injuries as more significant than a trolley scrape.

It thus seems unlikely that there is a simple, general formula that applies to actors in lesser evils justification cases. Sometimes the justification should be available if you do a lesser evil; sometimes it must be the least evil. In these sorts of cases then, the law will have to forgo the assurance Alexander would give that you will have a defense if you improve things. Instead the actor will know only that he will have a defense if his action is a morally sufficient improvement.

V. APPRECIATION OF JUSTIFICATORY SIGNIFICANCE

In earlier drafts of this paper I required for justification that the actor appreciate that the information would be accepted as justifying. Mike claimed such an appreciation in his last cross-examination response. Should we take him up on this point and insist, at least, that Mike think, as he runs the last few steps towards Vincent, "Well, now I will have a justification defense if I get arrested"? Is it different if he thinks, "I don't give a damn about any justification"? Jones and the stalker, without changing their motivations, could also have recognized that the new facts would give them justification defenses. I think the possibility of such a realization on any of their parts makes it clear how wrong it would be to convict Jones, Mike, or the stalker. Only the possibility of such a realization, however, not actual realization, should be necessary for the justification.

Suppose that Stan has determined to kill Moe, Curly, and Larry. Unbeknownst to Stan, the otherwise peaceable trio have unwittingly imbibed hallucinogenic punch. They burst through the door with axes and clear, if deranged, homicidal intent. They confront Stan just as he finishes loading his machine pistol. It happens, however, that Stan is under the misapprehension that the criminal law embraces the utilitarian principle that the number of innocent aggressors counts. He thinks that, although he would be justified in killing one

impaired innocent who is about to attack him, he would not be justified in killing three. So believing, Stan nonetheless guns down Moe, Curly, and Larry. He is justified as a matter of law, although he did not so much as appreciate that the circumstances were justifying.

What Stan could do at the time he fired was to describe the circumstances in a way that would be sufficient for us to conclude that he was justified, even though he thought he was not. Having a justification requires no more than reasonable belief in the existence of circumstances from which the justification can be inferred.³³ Just as ignorance of the law is, usually, no defense, ignorance of the law of defenses rarely, if ever, negates a criminal defense.³⁴

Although I think that these same considerations have application outside the criminal law with respect to moral justification, I do not think they apply there with the same force. Criminal defenses are almost always binary. They either apply, or they do not, with rare exceptions for jurisdictions recognizing mitigating, “imperfect justification.” In ordinary morality there is a good deal of room for partial justification and of justification for one purpose and not for another. There is also great flexibility in the informal sanctions of ordinary morality. To deny someone a moral justification, wholly or in part, because he or she was not motivated by the justifying reason will often work no significant injustice. This is not so when what is at stake is a criminal conviction.

VI. THE FORCE OF CONFESSIONS

It will not have escaped the reader’s attention that what few bad-motive justification cases come to light will almost all be confession cases. Are we simply to ignore the fact that these defendants readily confess that they acted out of hatred or lust rather than from any good reason?

We should ignore them because it would be a conceptual mistake to take these confessions as having legal significance for justification. Consider a Stan, who had nothing against his innocent assailants, gunning them down only because of the deadly threat they posed—but still with the false belief that he was committing murder. His confession that he had “just murdered three people,” would be relevant to whether he killed three people, but of no weight as to whether it was murder. Suppose an executioner blurted out that he hated the inmate so much that he would have injected him even had a stay come through at the last instant. We

³³ As shown by the following example, there will have to be some technical constraints on the inference. That the jigsaw puzzle, when put together, would spell out that Sam has a gun aimed beneath his robes would not justify shooting Sam even if you see all of the puzzle pieces and could infer the message if you had enough time.

³⁴ It would not negate the defense even if the defendant’s contemporaneous belief were that the actually justifying circumstances compounded the crime. Suppose a soldier carried out orders believing them to be unlawful and believing that carrying out unlawful orders would be an added specification at any court martial. He could turn out to be wrong on both accounts. His action could in fact be justified by the very superior orders he believed increased the criminality of his conduct.

would certainly want to remove him from his post, but we would not charge him with murder. Not even Mike can be sure whether he would have had a change of heart at the last second had he not seen that Vincent was about to attack Mary. He may know himself better than we do, but we will leave proof beyond a reasonable doubt far behind if we rest his conviction on his claim that he can properly unravel the key counterfactual. His speculations on the point should be deemed incompetent as a matter of law.

VII. CONCLUSION

Of the great majority of theorists who believe that awareness of justifying circumstances is necessary for justification, nearly all believe as well that the defendant must be in some respect motivated by those justifying circumstances. It should, however, be enough that the actor was aware of the justifying circumstances when she could still have changed her mind. As a practical matter, we would gain few additional convictions if we denied the defense of justification to bad-motive actors, because there are few cases in which it can be established that the actor was not taking justifying circumstances into account in some way. The few additional convictions would be gained at a cost of additional investigative intrusion.

As Alexander has argued, prosecuting those with awareness of justifying circumstances could create disincentives for actions that society generally wants to encourage. Those with knowledge of justifying circumstances, however, will not generally expect their bad motives to become known. So, the practical effects of the disincentives may be small.

The real reason not to convict someone on motive grounds who was aware of justifying circumstances is that it is unfair. It does not give the defendant proper credit for the fact that it was within her power to reconsider and, for example, decide not to save the stranger. Insofar as we would convict persons who have demonstrated dangerousness, these bad motive defendants have not so demonstrated. We do not, and cannot, have sufficient grounds for resolving the counterfactual that they would still have violated the law had they not been aware of the justifying circumstances.