



Nonviolent Protesters and Provocations to Violence

Shawn Kaplan

Abstract In this paper, I examine the ethics of nonviolent protest when a violent response is either foreseen or intended. One central concern is whether protesters, who foresee a violent response but persist, are provoking the violence and whether they are culpable for any eventual harms. A second concern is whether it is permissible to publicize the violent response for political advantage. I begin by distinguishing between two senses of the term *provoke*: a normative sense where a provocateur knowingly imposes an unjustified risk of a violent response, and a descriptive sense where the respondent feels provoked. I argue that, when the risk of a violent response is justified, the protesters are not provocateurs but akin to nonculpable, entrapping agents who create an opportunity for the regime to respond with disproportionate violence. The regime's response can disclose its brutality or criminality, and this can be fairly publicized for political advantage. When nonviolent protesters create an unjustified risk of violence either because the injustice they oppose is insufficient to justify the risk of a violent response or because the risked harms are disproportionate with the likely political advantages derived from the protest, they are partially culpable provocateurs. However, I argue these partially culpable provocateurs can still permissibly publicize the disproportionate violence of the regime so as to shape public opinion.

Keywords: nonviolent protest; provocation; risk imposition; self-defense; lawfare, *actio libera in causa*

I. Introduction

Though Gandhi proposed that the aim of nonviolent resistance is to allow one's opposition to become enlightened to the truth of their own injustice, the historical record shows that nonviolent protests are often met with disproportionate violence.¹ Whether one considers the brutal beatings by police of the marchers organized both by Gandhi in Dharasana and by Martin Luther King in Selma, or the deaths and injuries suffered by Palestinian protesters when soldiers opened fire, or the massacre of protesters in East Timor by the Indonesian military, the publicity from these disproportionately violent responses to protesters have significantly shaped global opinion regarding these conflicts.² In some instances, the violent responses were foreseen, yet the protesters persisted. This has led to some nonviolent protests being labeled as provocations to violence. Michael Gross has argued in *The Ethics of Insurgency* that to engage in an activity when one reasonably foresees a violent

response is sufficient to label the initial action a provocation—regardless of the agents' intentions or the justice of their cause.³ For Gross, the role of provocation is central for analyzing the ethics of nonviolent resistance, especially by insurgents.

There is something troubling about Gross' claims regarding provocations to violence. It seems that there are many heroic or praiseworthy actions that would satisfy his account of a provocation, e.g., the person who steps between a bully and his intended victim. This person does not intend to be the target of violence but reasonably foresees it to be an outcome of his actions. In addition, we can imagine cases where innocent individuals engage in permissible actions for which they can reasonably foresee a violent response. Consider the student who has been threatened with a beating by a bully if she attends school the next day but decides to attend nonetheless. Also, a homeowner who confronts a domestic intruder can also reasonably foresee a violent response by the intruder. Both the threatened student and the homeowner are within their rights to attend school and defend against an intruder respectively. We can perhaps debate the prudence of all three decisions, but labeling them provocations seems to unfairly shift culpability for the violent response onto the praiseworthy or innocent individuals.

Can the same be said of nonviolent protesters who persist regardless of foreseeing a violent response? Consider the following case:

Nonviolent Marchers (NM): *In response to policies of a colonial government which unjustly limits access to a necessary resource, a non-violent march is organized to gain access to this resource. Anticipating that their way will be blocked by police, organizers of the march instruct participants that they should not disperse when so ordered and that they should not resort to violence, even if they are attacked by the police. The potential for police violence is foreseen but it is not the intended aim of those who march or their organizers.*⁴

If the marchers are met with police violence after they are ordered to disperse and are issued a warning, have they provoked this violence?

To answer this question, we ought to first recognize that NM differs from the three previous examples. In the case of the person heroically intervening on behalf of the individual being bullied, the intended aim is to stop the bullying and the foreseen risk of violence is a negative side-effect. Similarly, for both the threatened student and the defensive homeowner, the reasonably foreseen violence that may result from their actions is a side-effect of their permissible aims. In NM, there are two significant differences. First, the marchers are engaged in civil disobedience wherein the law is being broken and, thus, they are in need of a justification for their actions in a way the other examples do not require. However, even if the civil disobedience of the marchers is unjustified, neither the police nor any bystanders are wronged by the nonviolent protesters in a way that would justify defensive force by the police. By not complying with the order to disperse, the marchers will, at worst, commit a wrong against the rule of law and this wrong would be distributed broadly across all members of society

and, thus, would not make the protesters liable to defensive force. Second, while the foreseen violence is not the intended aim of the marchers or their organizers, it can be used instrumentally to shape public opinion. For example, the brutality of the police beatings of the marchers organized by Gandhi at Dharasana in 1930 received global attention originating from a single newspaper article. It is widely held that this publicized brutality helped to turn global public opinion against British rule in India and in favor of the liberation movement. More generally, publicizing disproportionate violence used against protesters (especially nonviolent protesters) can have a strategic effect, suggesting that the foreseen violence is not *merely* a negative side-effect of the action. Does this then mean that the marchers have provoked the violence? This seems to be the case for Gross as he emphasizes the strategic role of publicizing disproportionate violence against protesters.⁵

At this point, it will be helpful to unpack what might be entailed by labeling a nonviolent protest a provocation to violence. Provocation is a thorny concept that is sometimes employed descriptively but often normatively. Gross is not especially clear on how he understands the nature of provocations. At various points, he uses the following phrases as synonyms for provoking violence: precipitating violence, soliciting violence, and inciting violence. To my mind, these phrases are not equivalent expressions. One could unknowingly precipitate a violent response, e.g., by unwittingly carrying a red flag near a bull. In contrast, one cannot solicit or incite violence unknowingly. In addition, soliciting a violent response, e.g., by daring another person to punch you, can differ from inciting violence, e.g., by stirring up a powerful emotional response. Some may argue that one could unwittingly stir up the emotions of another and, thus, unknowingly incite violence. However, it is important to keep in mind the difference between a person feeling provoked and an agent acting as a provocateur. Kimberly Ferzan draws out this contrast quite well:

An act can incite someone toward violence, even if the person who committed that act was not aware of its inciting properties. For instance, imagine that Ed is sleeping with Sally, but does not know, nor does he have any reason to know, that Sally is married. If Sally's husband Stan finds the two in bed together and shoots Ed, Ed's actions certainly provoked Stan's response. However, Ed, it seems, should not count as a provocateur as he was unaware of the fact that his actions could even have that effect.⁶

By suggesting that Ed's actions unknowingly provoked Stan's violent response, Ferzan is giving a description of how Stan's mental state was affected by Ed's action. This description of the factual state of affairs does not affect how we normatively assess Ed's action. He has not acted as a provocateur towards Stan, but his actions have unwittingly contributed to a situation where Stan *feels* provoked. As such, Ed does not hold any moral responsibility for Stan's violent response.⁷ Following Ferzan's account, to be a provocateur one must be aware that one's actions could provoke

another person. More specifically, she claims provocateurs knowingly create “the unjustifiable risk of causing the respondent to attack him.”⁸

In assessing the ethics of nonviolent resistance, we ought to be primarily concerned as to whether the protesters have acted as provocateurs, as opposed to whether the state authorities feel provoked. That is, we ought to consider whether protesters provoke in the agential sense of the verb by knowingly creating an unjustified risk of a violent response. Whether the opposition is provoked in the passive, non-agential sense of the verb is only relevant when considering if their culpability for the violent response is partially mitigated. Ferzan presents two ways that a provocateur can consciously create an unjustified risk. First, a risk is unjustified when the action itself is either impermissible or produces the risk of disproportionate harm. Returning to some of our previous examples, the threatened student and the defensive homeowner are not provocateurs insofar as they are within their rights to attend school and defend against a domestic intruder *but only if* the risks created do not threaten disproportionate harm. A second way to consciously create an unjustified risk arises when the societal role of the individual does not permit her to create the risk.⁹ Ferzan has in mind instances where the provocateur is doing something analogous to entrapment. Since the individual is not authorized to entrap wrongdoers (like the police might be), she claims that person is consciously creating an unjustified risk of criminal or violent behavior. This conclusion, however, is contingent upon “the existence of a well-functioning state.”¹⁰ This part of Ferzan’s analysis will prove useful for considering the case of NM. Like the entrapping agent who “creates an opportunity” for another to act criminally or violently, perhaps our marchers simply create an opportunity for the police to respond with disproportionate violence?

In what follows, I first develop in greater detail Gross’ arguments regarding provocations to violence by those engaged in nonviolent resistance. By focusing on the distinction between an agent being a provocateur and a respondent feeling provoked, I expose some confusions in Gross’ claims regarding the justice of nonviolent resistance by insurgents. Only provocation in the agential sense is relevant for assessing the ethics of nonviolent protest. I argue that, while nonviolent protesters hold a share of the moral responsibility for the reasonably foreseen risk of violence they help create, they can act as non-culpable, entrapping agents when that risk of violence is justified. In addition, I explore how the contrast between fair and unfair entrapment can help assess the ethics of nonviolent resistance. In Section III, I examine the *actio libera in causa* legal debate that asks whether provocateurs can justifiably defend themselves after causing the conditions of their own self-defense. Using arguments from this debate, I show that provocateurs who intend to produce unjustified risks of violence from their nonviolent protests can sometimes still be justified in taking advantage of the situation they have wrongly created in order to negatively shape public opinion regarding their opposition.

II. Nonviolent Resistance as Provocation or Entrapment

In *The Ethics of Insurgency*, Michael Gross seeks principles for justifying nonviolent resistance when deployed by the political wing of insurgent movements. Controversially, he proposes a revised set of just war principles for assessing the justice of nonviolent resistance against the state, equating such tactics with “soft war.” He assumes that an insurgency’s commitment to nonviolent resistance is “pragmatic as opposed to principled” and that nonviolent resistance should not be confused with pacifism.

Nonviolent resistance is not pacifism—that is, a moral repugnance of violence—but a strategic choice when the costs of waging an armed resistance are high and its chances of success against a militarily superior adversary are low. Nonviolent resistance can be no less coercive than armed violence is. Its goal is to undermine the prevailing asymmetry of power and force a state to yield to an aggrieved group’s demands by threatening the state’s material interests and international image.¹¹

While Gross acknowledges that nonviolent resistance can have the aims of promoting solidarity and self-help within the aggrieved group, he focuses upon its “coercive” aims against the state.¹² According to Gross, successful nonviolent resistance creates a sort of practical dilemma (as opposed to a moral dilemma) for the regime in power. The regime either concedes political ground to the protesters or acts in a way that undermines its own legitimacy, e.g., by resorting to disproportionate violence against the nonviolent resisters. It is here that Gross sees the problem of provocation emerge.

Especially vexing is the question of provocation. Many forms of nonviolent resistance depend on “backfire,” that is, a harsh, disproportionate, and violent state response to nonviolent protest that may put nonviolent protesters at considerable risk. This risk, often foreseen by organizers, is analogous to the danger guerillas can expect when employing human shields.¹³

Backfire can be effective in altering global public opinion and, thus, can potentially help the political cause of the aggrieved party by gaining various forms of international support (e.g., verbal or economic support, the imposition of sanctions, or military support/intervention). In addition, by publicizing the disproportionate violence of the state, divestiture and boycott campaigns can gain momentum so as to inflict greater material damage to the state’s economy.

So long as the response is violent, Gross views the nonviolent resistance to be a provocation. He indicates that there are at least three types of provocation to

violence: acts where the respondent's violence is neither intended nor reasonably foreseen, acts where the respondent's violence is both intended and foreseen, and acts where the respondent's violence is not intended but reasonably foreseen.¹⁴ That an unintended and reasonably unforeseen violent response is sufficient for Gross to label the nonviolent resistance a provocation can only indicate that he is being descriptive of the respondent's mental state, where those responding *feel* provoked by the protest. Insofar as the violence would not have occurred without the nonviolent protest, we can fairly describe the protest as precipitating the emotional response, but we cannot describe the protesters or their organizers as provocateurs. If the violence is reasonably unforeseen, then it is impossible to shift any moral responsibility for the violent response and resulting harms onto the protesters and organizers—as one might if they were aware of creating a risk. In contrast, moral responsibility for an intended *and* foreseen violent response ought to be partially held by the nonviolent resisters; moral responsibility is shared with the violently responding police and, possibly, with those officials who created the policy that foreseeably led to the protests.¹⁵ Gross is correct to point out that the partial moral responsibility of the nonviolent resisters for knowingly and intentionally putting protesters at risk does not automatically make such nonviolent resistance unjustified.

Answering this question calls to mind the calculations of political leaders as they risk their compatriots' lives during war. Protesters and demonstrators are foot soldiers, and the criteria for a legitimate nonviolent struggle are the same as for armed struggle: just cause, legitimate authority, a reasonable chance of success, equitable conscription, consent, and proportionate human costs in view of the goals insurgents seek.¹⁶

At least in the context of a full-scale insurgency, Gross' analogy between nonviolent and violent resisters is not altogether farfetched.¹⁷ I am not going to detail the ways Gross revises the just war conditions of just cause, proper authority, and reasonable chance of success to better accommodate insurgencies as this will take us too far afield. Let us, for now, broadly affirm that the injustice being opposed by the resistance must be great enough to justify the foreseen and intended risks created and that the likely benefits of creating the risk must outweigh the likely harms to innocent individuals. In addition, I will follow Gross' claim that organizers of nonviolent resistances are prohibited from coercively conscripting participants and that protesters must be fully informed of the risks so as to give informed consent. If all the justifying conditions are satisfied when a violent response is both intended and foreseen, is this *truly* a provocation where this term has some normative significance and isn't merely descriptive of what the respondent feels?

Though it might be quite rare that organizers (or protesters) are justified to intentionally and foreseeably create a risk of violence to protesters, such a justification would establish a liberty right and, thus, is analogous to our earlier examples of the threatened student and defensive homeowner.¹⁸ Since the risk they impose is justified,

it's difficult to see how the protesters or their organizers hold any moral liability or culpability for the risk of violence or any consequent harm. Liability, at minimum, requires that one is morally responsible for posing an *unjustified* risk of harm onto others. Insofar as the protesters are conscious of a foreseeably violent response, they hold partial moral responsibility for the risk created. However, insofar as this risk is justified, the protesters have not wronged the police and they are neither liable to defensive harms nor for compensation to those who are harmed. If we follow Jeff McMahan's account of moral liability, then we should expect there to be some inevitable harm or cost which ought to be distributed to the party who holds the greatest moral responsibility for that inevitable harm or cost so as to protect those who are less responsible.¹⁹ Liability is not equivalent to desert, or what people are owed, but what is the most just distribution of the inevitable harm or cost.²⁰ However, in the scenario where the protesters and their organizers not only foresee a violent response but intend it, they have *not* created a situation where harms are inevitable and require just distribution. Instead, they have contributed to a foreseen and intended risk of violence but, for that risk to come to fruition, the police must choose to attack the protesters. Though Gross casts the protesters as creating a dilemma for the police and authorities between two undesirable outcomes (conceding political ground or conceding political legitimacy by attacking the protesters), he ignores that the police could simply attempt to round up the nonviolent protesters for arrest or engage in a prolonged stand-off. Since the harms only become "inevitable" once the police attack, liability falls squarely on them for their intentional and disproportionate violence.²¹ Furthermore, by presenting the police with a choice where their brutality can be displayed, the protesters are not in any way culpable for the violence—just as the police are not culpable when they offer illegal choices and entrap citizens (e.g., when a police officer poses as a drug dealer). At the same time, if the protesters are engaged in something analogous to entrapment, then we need to consider whether this entrapment is somehow unfair or unjust, as sometimes can be the case with police entrapment.

Ferzan suggests that "one concern underlying entrapment is about external actors pulling the strings on our circumstantial luck."²² While the police may be permitted to modestly alter our circumstantial luck, we should draw limits on more elaborate manipulations. For example, some police in the U.S. will entrap people who are vulnerable due to poverty or mental illness by offering a large incentive to carry out a criminal act. Their target would likely never have considered the criminal act otherwise but the police convince them to act criminally either with the promise of a large payment or by preying on their mental instability. Compare this to a case where the police pose as drug dealers and the drug buyer simply has the bad luck of approaching the undercover officer. This second form of entrapment only modestly alters the circumstantial luck of the respondent and few complain that the unlucky have been treated unfairly in the process. However, the more elaborate entrapment scheme that creates novel reasons for criminal activity may strike one as unfair for the very reason that it is manipulating people to commit crimes they would not otherwise consider.

Where does the case of the nonviolent protesters who refuse to disperse when ordered and warned stand in comparison to these two types of entrapment? One might argue that the police blocking the marchers' progress would not have sought out to brutally beat citizens if not for the protesters' refusal to obey their commands to disperse. However, it is more difficult to claim that the nonviolent protesters have manipulated the situation to create novel reasons for a brutal response. Given that the police respond with brutality and not with either an attempt to arrest the protesters with proportionate force or a concession of political ground to the protesters, it may be reasonable to infer that their choice discloses to the court of global public opinion the injustice and the, up until then, latent brutality of the regime. If any nonviolent challenge to the regime's authority is likely to be met with brutality, then protesters seem to only modestly alter the circumstantial luck of the police. Thus, those responding with disproportionate violence to the protesters cannot readily claim that they have been unfairly entrapped.

However, Ferzan raises a red flag regarding entrapment carried out by everyday citizens. She uses the example of the character played by Charles Bronson in the 1974 film *Death Wish* to illustrate her concern.²³ Bronson's character poses as a weak and vulnerable person late at night in 1970s New York City hoping to draw criminals to attack him so he could kill them in self-defense.²⁴ Like the police officer posing as a drug dealer, Bronson only modestly alters the circumstantial luck of his eventual attackers, since he is not creating novel reasons for them to resort to criminal violence. At the same time, there is an important difference for Ferzan insofar as Bronson is acting as a vigilante and lacks the authority to entrap wrongdoers. Recall that this is for Ferzan the second way one can consciously create an *unjustifiable* risk of a violent or criminal response. Here, walking late at night in the dangerous streets and subways of New York is something Bronson has a liberty right to do but Bronson's societal role does not permit him to entrap these would-be criminals. This limit that "fair entrapment" must be carried out by those with social authority only holds, according to Ferzan, when operating within a well-functioning state.²⁵ A case could be made that those living under occupation by a colonial power do not live within a "well-functioning state," especially when—as in NM—the regime is unjustly limiting access to needed resources. If this is so, our protesters who foresee and intend to create a justified risk of violent response from the agents of the colonial regime do so under the heading of "fair entrapment."

We have been analyzing a scenario where the protesters both foresee and intend for the police to respond with violence for the sake of trying their case in the court of public opinion. We have also stipulated that the creation of this risk is justified. Though the police authorities may *feel* provoked by the nonviolent protesters who confront them and refuse to disperse, we have followed Ferzan's point that the protesters must create *unjustified* risks to be fairly cast as provocateurs who share in the culpability for the violence and consequent harms. The risk created by the protesters would be unjustified if their grievances against the state are not significant enough to provide a

just cause for putting the protesters at serious risk of harm, or if those likely harms would be disproportionate to the potential benefits of the negative publicity for the state. When those risks are justified and the protesters do not live within a well-functioning state, we have argued that their behavior is analogous to “fair entrapment” where the circumstantial luck of the state authorities is only modestly altered. However, this differs in a key way from NM where the protesters foresee but do *not* intend the violent response. Entrappers offer the choice for others to respond violently or criminally with the intention of exposing the brutality or criminality of those who make that choice. While those who create the risk of foreseen but unintended violence with their protests are not engaged in a perfect equivalent to entrapment, the foregoing analysis still remains relevant.

The nonviolent protesters in NM who foresee but don’t intend a violent response still offer a choice to the police and the colonial authorities. The intended aim of the nonviolent protesters doesn’t change the fact that the police can choose to either: concede political ground, wait out the protesters, use proportionate force to arrest the protesters, or attack them. The choice they make could disclose much about themselves and their regime. This is true regardless of whether the nonviolent protesters specifically intend a violent response or hold the more general intention of altering the power asymmetry between themselves and the regime they oppose by whatever means their opposition chooses. (The protesters can potentially gain success in altering the power asymmetry either: by gaining immediate concessions, by publicizing their cause through either a lengthy stand-off with the police or by filling the court system after mass incarceration, or—if attacked—by publicizing the regime’s brutality.) Though it would be odd to call this entrapment, since the violent response is unintended, it shares with entrapment a similar structure. At T_1 , an initial action poses a choice to the opposition and at T_2 , after the opposition’s criminal or brutal response, the initial actor is permitted to take advantage of that response so long as they have not significantly altered the circumstantial luck of the respondents. Thus, for our original NM case, we can conclude that the protesters can permissibly publicize the disproportionate force of their adversaries so as to damage their international reputation and alter public opinion in their own favor.

So far we have only considered cases where the initial actors do not share in the culpability for the violent response because the risk they create is justified. In the next section, we will consider scenarios where the actors are provocateurs because they consciously create an unjustified risk at T_1 , and we will consider what this means for the permissibility of their taking advantage of the violent response at T_2 . We will analyze these questions by considering the legal debate of *actio libera in causa* where one causes the conditions for one’s own self-defense.

III. Provocateurs and *Actio Libera in Causa*

Provocateurs consciously create an unjustified risk of a violent response at T_1 . Should this alter how we morally assess the permissibility of them taking advantage of that violent response at T_2 ? Or should the permissibility of and culpability for the two actions be assessed completely separately from one another? These two alternatives represent opposing sides within the *actio libera in causa* debate.²⁶ In this section, we will examine these opposing positions in the context of creating the conditions for one's own self-defense and then we will consider how this relates to creating the conditions for generating negative publicity from backfire.

Let's consider the following to reflect upon cases of *actio libera in causa*.

Taunter: *At T_1 a provocateur taunts his innocent target so as to consciously create an unjustified risk of a violent response. The risk is unjustified because the target has done nothing to warrant the provocation and the produced risk is not a lesser evil of some sort. Is the provocateur justified in using proportionate and necessary force to defend himself from the provoked response at T_2 ?*

This is a clear case of creating the conditions for one's own self-defense. Though the target has been provoked, this does not mean that the violent response to the provocation is justified. The taunting done by the provocateur does not rise to the level of aggression and, thus, the violent response is not defensive. At most, some of the respondent's culpability may be mitigated. However, Ferzan argues that the provocateur is similar to an aggressor:

Provocateurs forfeit defensive rights for the very simple reason that provocateurs start the fight. They initiate fights without initiating aggression. They are actually stealth aggressors. Although initial aggressors start fights by throwing first punches, stealth aggressors start fights by inciting the other person to anger or otherwise causing the other person to do violence.²⁷

Ferzan suggests that provocateurs and aggressors are similar insofar as both "start the fight." Aggressors can be understood to forfeit their right to self-defense insofar as they are morally responsible for posing an objectively unjustified threat of harm to another.²⁸ When a person defends herself against an aggressor, the aggressor has no moral complaint against the defensive harms inflicted—so long as the defense is proportionate and necessary. Ferzan claims that the provocateur similarly has no moral complaint against the person he provoked. "Having created this risk of harm to himself, he forfeits his moral complaint when this very risk materializes. It is true that this conduct is insufficient to justify the respondent's behavior, but sometimes, there are just two wrongdoers."²⁹ The key for Ferzan is that the provocateur is *culpable* at T_1 for consciously creating an unjustified risk of harm. This culpability at T_1 is sufficient, according to Ferzan, to alter the moral relationship between the provocateur

and target at T_2 such that the provocateur has forfeited his right to self-defense. It is worth noting the specificity of this forfeiture of defensive rights. The provocateur has only forfeited this right to his target but maintains defensive rights against all other parties.

Ferzan's argument that both the respondent's violence and the provocateur's defense are both unjustified is problematic. I agree that, by consciously creating an unjustified risk of a violent response, the provocateur holds culpability for that risk and that this culpability is not sufficient to justify the respondent's violence. However, if this culpability is sufficient for the provocateur to forfeit his right to self-defense, as Ferzan claims, who is wronged by the violent response? Clearly for Ferzan, the provocateur is not wronged, as he has "no moral complaint" after having "started it." How is the respondent a wrong-doer, if no person is wronged by her violence? Perhaps Ferzan might respond that there is a wrong committed to the general public or to the rule of law, but this seems a bit of a stretch given the unjustified violence inflicted upon the provocateur. Second, I find calling the provocateur a "stealth aggressor" because he "started it" worrisome. This sort of playground rationale that the taunter "started it" and, thus, has no moral complaint when met with the respondent's violence is far from convincing. We can fairly claim that the provocateur should not have taunted the target (since an unjustified risk was knowingly created) and that the target should not have taken the bait and responded violently. However, it is less clear that the provocateur must permit himself to be violently harmed because the respondent *wrongly* took the bait. If the baiting and taunting were extreme, one might argue that the culpability for the violent response is partially mitigated. At the same time, heightening the rhetoric by calling the provocation "stealth aggression" is not enough to conclude that the provocateur is liable to those harms. Unless one can show that the provocateur has *greater* moral responsibility, or *greater* culpability, for the violence than the respondent, then one cannot claim he is liable to those harms. Ferzan does not make this case, nor do I see how one could. Returning to McMahan's account of moral liability, we require a case where a cost or harm is inevitable and needs just distribution. However, the taunting provocations do not make the violent response inevitable and the provocateur is not more culpable for the violence than the respondent who acts freely. Since the provocateur is not liable to defensive harms, how can we conclude that he forfeited his right to self-defense? Lastly, to suggest that the provocateur "had it coming" as he started the fight with his taunts is to affirm a brand of "playground justice" that is ripe for abuse.

In contrast, Larry Alexander argues that we should separately assess the actions at T_1 and T_2 .³⁰ He expresses his basic position most succinctly:

If one assesses the culpability of an actor at each of the various times he acts in a course of conduct, then it is obvious that he can be nonculpable at T_2 but culpable at T_1 , and that a nonculpable act at T_2 has no bearing on whether an actor was culpable at T_1 when he caused the circumstances that are exculpatory with respect to his act (or conduct) at T_2 .³¹

In the case of the provocateur who taunts her target into responding violently, Alexander agrees with Ferzan that the provocateur is culpable for creating an unjustified risk of violence at T_1 . However, he rejects that this causes the provocateur to forfeit her right to self-defense at T_2 . He argues that the respondents who have taken the bait have become “culpable aggressors whose attacks are neither justified nor excused by the provocations.”³² He avoids the problems associated with Ferzan’s position which requires the provocateur to suffer unjustified violence by arguing that the initial culpability can justify the provocateur’s later judicial punishment but not a forfeiture of the right to self-defense. He supports his position by asking us to imagine a third party who is present during the provocation and the violent response. Alexander asserts that the third party onlooker would be justified in defending the provocateur from the respondent’s attack because, even if the provocateur is culpable because of the taunting, the person who attacks has greater culpability as an aggressor. He reasons that, if a third party is justified to defend the provocateur, so too is the provocateur herself. “In other words, because if they [the provocateurs] are culpable at all, they are less culpable than their attackers, they are justified at T_2 in defending themselves for the same reason that a third party, apprised of the plots, would be justified in defending them—namely, their lesser culpability than that of their attackers.”³³ Alexander answers our earlier question, directed towards Ferzan, comparing the culpability of the taunting provocateur and the respondent. Culpable provocation will always be outweighed by culpable aggression and, thus, for Alexander the violent respondent will be liable to defensive harms.

I agree with Alexander that the taunting provocateur, who both foresees and intends to create a risk of a violent response, can be held morally and legally culpable for that increased risk at T_1 but that this can be kept separate from the question of whether the provocateur’s self-defense is justified at T_2 . Since the respondent has *wrongly* taken the bait and attacked, there is no reason for why the provocateur must allow herself to be subjected to these unjustified harms. Culpably “starting it” with provocative fighting words pales in comparison to culpably aggressing against the provocateur. Towards the end of his essay, Alexander suggests that the intentional provocation by the taunter serves as a sort of oblique solicitation to attack. This makes the provocateur complicit in the attack, but that complicity will always make for a lesser degree of liability than what the actual aggressor holds.³⁴

How might Alexander’s analysis of *actio libera in causa* illuminate a case of nonviolent protesters consciously creating an unjustified risk of a violent response? To keep the cases as similar as possible, let’s consider a case where the protesters verbally taunt the police and by doing so create a novel reason for a violent response. The risk they create is both foreseen and intended with an ambition to publicize the police violence so as to shape public opinion. We are also assuming that the risk is unjustified, either because the injustice the protesters are combating is not sufficient to justify the risk or the potential benefits from shaping public opinion will be outweighed by the probable harms from the police violence. Though the risk of

violence the protesters create is unjustified, this does not justify or excuse the police taking the bait and responding violently. As provocateurs, Alexander would argue they are culpable for the risk they have created and, thus, deserve moral censure and, possibly, criminal punishment regardless of whether the police respond violently. However, the protesters' culpability at T_1 will be outweighed by the culpability of the police who are culpable aggressors at T_2 . Thus, the protesters do not forfeit their right to self-defense at T_2 . Since third parties would be justified in protecting the provocateurs from the police violence, so too the protesters would be justified to defend themselves as well. Now, unlike the question of contrived self-defense that is taken up by Alexander and the other figures debating *actio libera in causa*, we are asking whether the provocateurs would be justified in taking advantage of the police or state authorities responding violently, e.g., by posting videos of the violence on the internet so as to shape global public opinion. Given the greater culpability of the violently responding police and state authorities, they have no moral complaint for having their deeds tried in the court of public opinion.³⁵ Since the tactic of backfire is less harmful than taking up arms in self-defense against the state, it seems reasonable that the standard for permitting it would be less restrictive than that for permitting contrived self-defense. If Alexander's argument is successful in justifying self-defense by culpable provocateurs, surely it will justify the less harmful tactic of publicizing the respondent's disproportionate violence so as to alter public opinion and shift the power asymmetry between the state and the insurgency.

It is important to note that this argument might not hold if the provocation were more significant than verbal taunts. What if the protesters threw rocks at the police, or were engaged in the destruction of property, or were brandishing weapons? Let's assume that the aim of these other provocations was still to elicit a disproportionately violent response in the hope of shaping public opinion. In the most severe case of brandishing weapons, would a third party be justified in defending the armed protesters from the police's violent response? It is conceivable that, even if the armed protesters have no intent to use their weapons in an act of aggression, they can still be liable to defensive harms from the police insofar as they are morally responsible for appearing to be an objectively unjust threat. Jeff McMahan argues that if someone is responsible for the appearance of being an objectively unjust threat, and thus responsible for the reasonable belief that defensive harms are necessary to stop him from threatening innocent people, then this person cannot complain of being wronged if targeted with proportionate defensive force—nor would this person or a third party have the right to defend against this attack which is merely grounded in an evidence-relative permission.³⁶ Two key caveats in McMahan's position are that these individuals must be morally responsible for the *reasonable belief* that defensive force is *necessary*. If we interpret necessity to indicate that there is no less harmful means of averting unjust harms, then we might be reasonably dubious that merely brandishing weapons (e.g., baseball bats) is sufficient to trigger the necessity condition for McMahan's account of liability. If the police can reasonably believe that they can use less harmful means (e.g., waiting out the armed protesters) to avert unjust harms, then McMahan's caveats

undercut the claim that the protesters are liable to defensive harms and that the police have an evidence-relative permission to use proportionate force.

There may, however, be another and less contentious argument that can eliminate the permission of the provocateurs to take advantage of the state's disproportionate violence at T_2 for cases where the protesters either brandish weapons or destroy property. Publicizing police violence against protesters who are engaged in either of these two more extreme acts of provocation is unlikely to sway public opinion in favor of the protesters. Even if many in the public perceived the police response as being disproportionate, they would likely not become more sympathetic to the protesters' cause when they were holding weapons or destroying property (especially if this resembled a riot). Even if the police are culpable for using disproportionate violence, the public perception will be that the armed or destructive provocateurs held greater culpability. At bottom, the tactic must have some reasonable likelihood of success in shaping public opinion and this condition will not likely be satisfied so long as the protesters are violent, destructive, or threaten violence.

The case of the protesters throwing rocks is less straightforward. We have seen in the case of Palestine that images of protesters throwing rocks at heavily armed security forces have been effective in shaping public opinion against the Israeli occupation—especially when those protesters have been children. This has been true even when the security forces have not responded violently to the provocation. The tactic of throwing rocks to provoke a disproportionately violent response cannot be so quickly ruled out by a likelihood of success principle. At the same time, though the rock-throwing protesters are morally responsible for posing an unjust threat of harm to the police and are liable to defensive harms, the police are required to use proportionate force to disperse or arrest the protesters. Like the case of the verbal taunters, if the rock throwers provoke a disproportionate response, then a third party would be justified in defending them from the attack. Their provocation at T_1 would not result in the forfeiture of their right to self-defense from disproportionate force at T_2 and they would be permitted to publicize the disproportionate response from authorities to shape public opinion. Unlike the case of the verbal taunters, the rock throwers would be liable to proportionate defensive force to avert the harm from the rocks.³⁷ Like the taunters, they would have no moral complaint against being dispersed or arrested by means of proportionate force and would have no clear justification for taking advantage of the proportionate response at T_2 in an attempt to shape public opinion.

In this section, we have examined cases where protesters are provocateurs because they consciously create an unjustified risk of a violent response. Unlike the previous cases of "fair entrapment," the protesters create novel reasons for the violent response but this does not justify or excuse this violence. The *actio libera in causa* debate has helped us to analyze the ethics of provocateurs taking advantage of the disproportionately violent response they have provoked in order to shape public opinion. By siding with Alexander's analysis, where we appraise the culpability for the provocation at

T_1 and the taking advantage of the disproportionate response at T_2 separately, we have found cases where the foreseen and intended unjustified risks created by the provocateurs do not undermine a permission to take advantage of the unjustified and unexcused responses. The permission to publicize the disproportionate responses so as to shape global opinion is limited by a reasonable likelihood of success condition which is unlikely to be met when the provocations involve violence, destruction of property, or credible threats of violence.

IV. Concluding Comments

In recent years, groups engaged in nonviolent resistance have employed tactics that put their own populations at serious risk of harm. Often enough, the potential harms to their own population are foreseen and sometimes appear to be an intentional aim of the tactic. I have argued that nonviolent protesters hold a share of moral responsibility for the risk, so long as they are conscious of creating that risk. If the injustice they are resisting is severe enough to justify the risk and the benefits produced by publicizing the disproportionate harms outweighs those harms, then the resisters are not acting as provocateurs. Instead, I have argued that they act analogously to the police who engage in “fair entrapment” by offering their opposition a choice whereby they can disclose their brutality or lack thereof. On the other hand, protesters are provocateurs when they consciously create an *un*justified risk of a violent response. Provocateurs hold partial culpability for any harms suffered as a result of their unjustified risks. However, the fact that provocateurs hold partial culpability for the unjustified risks they create and for the resulting harms to their population neither justifies nor excuses their opposition taking the bait and engaging in a disproportionate attack. I have argued that the relative culpability of the disproportionate respondent far outweighs that of the provocateurs and, thus, makes them liable to defensive harms from *anyone* who could avert the unjustified harms to the protesters. Using a similar logic, I have concluded that the provocateurs can permissibly take advantage by publicizing the disproportionate attack at T_2 , even though they intentionally created the unjustified risk of that attack at T_1 . This permission is limited by a likelihood of success condition which is not readily satisfied if the provocateurs are either violent, destructive, or pose a credible threat of violence.

Shawn Kaplan is an Associate Professor of Philosophy and the Director of the Ethics & Public Policy Program at Adelphi University in New York. His chief areas of research are: the ethics of war and armed conflict, just war theory, international ethics, and Kantian ethics. In addition to articles on Kantian ethics and just war theory, his publications include a series of journal articles on the nature of terrorism and its potential justifications. More recently, his research has focused upon: the ethics of counterterrorism tactics, the ethics of nonviolent resistance, and problems of privacy in regard to facial recognition surveillance. He is currently working on a theory of just armed conflict under conditions of uncertainty.

Notes and References

1. Mahatma Gandhi, "Ahimsa or the Way of Non-violence," in *All Men Are Brothers: Life and Thoughts of Mahatma Gandhi* (Paris: UNESCO, 1958), 85-107.
2. In the case of Palestine, we tend to see an inconsistent application of nonviolent tactics, as was witnessed in 2018. Regardless, the seemingly disproportionate response by Israel has caused significant outcry and has helped fuel the divestiture and boycott movement.
3. Michael L. Gross, *The Ethics of Insurgency: A Critical Guide to Just Guerilla Warfare* (Cambridge: Cambridge University Press, 2015), 253.
4. This case highlights salient features of Gandhi's Dharasana Salt Works March in 1930.
5. Gross, *The Ethics of Insurgency*, 244-5.
6. Kimberly Kessler Ferzan, "Provocateurs," *Criminal Law and Philosophy* 7, no. 3 (2013): 597-622, at 610.
7. Moral responsibility requires one to either act knowingly or that one's ignorance is due to one's own negligence.
8. Ferzan, "Provocateurs," 600.
9. Ibid.
10. Ibid, 620-1.
11. Gross, *The Ethics of Insurgency*, 242.
12. Depending upon what theory of coercion one adopts, this description will make more or less sense. Gross seems to assume the view that any action that forces someone to give up her preference is coercive. Critics of this view claim it captures too many non-coercive acts. For example, this interpretation suggests that, if I arrive early to be first in line, then I coerce all those whose preference it was to be first to cue up behind me. Moral interpretations of coercion, in contrast, require that one would need to be forced to give up an action one has a right to choose. In this second interpretation, coercion is a *prima facie* wrong. Pressuring an oppressive regime to properly respect the rights of the oppressed party wouldn't be coercion from this perspective.
13. Gross, *The Ethics of Insurgency*, 241.
14. Ibid, 253.
15. The moral responsibility for the violent response may be shared unequally between these parties but the exact proportions of responsibility will be determined on a case-by-case basis. Such an analysis exceeds the bounds of the current argument.
16. Gross, *The Ethics of Insurgency*, 254.

17. I am less comfortable drawing the analogy when the nonviolent resistance is a tactic within a more focused campaign of societal reform.

18. The analogy is imperfect, since in those previous cases the foreseen violence was not intended. However, the fact that the risk was intended—perhaps to trigger negative publicity—does not by itself necessarily make the nonviolent protest unjustified and, for the sake of argument, we have stipulated that this risk is justified.

19. Jeff McMahan, “Justification and Liability in War,” *The Journal of Political Philosophy* 16, no. 2 (2008): 227-244.

20. Jeff McMahan, *Killing in War* (Oxford: Oxford University Press, 2009), 8.

21. In contrast, if the protesters were violent and attacked the police, then there would be inevitable harms either to the police and the protesters. Following McMahan’s theory of moral liability, violent protesters would be liable to defensive harms due to their greater moral responsibility for the inevitable harms.

22. Ferzan, “Provocateurs,” 620.

23. Ibid, 620-1.

24. *Death Wish* (Paramount Pictures, 1974).

25. Ibid.

26. A third alternative is presented by Finkelstein and Katz where the way we assess the permissibility of the act at T_2 should alter how we assess the initial provocation. Since they do not present this position to be their definitive stance and it is plagued by unsupported assumptions, I will not take it up here. Claire Finkelstein and Leo Katz, “Contrived defenses and deterrent threats: Two facets of one problem,” *Ohio State Journal of Criminal Law* 5, Spring (2008): 479-504.

27. Ferzan, “Provocateurs,” 615.

28. McMahan, *Killing in War*, 34-5.

29. Ferzan, “Provocateurs,” 615.

30. Larry Alexander, “Causing the Conditions of One’s Defense: A Theoretical Non-problem,” *Criminal Law and Philosophy* 7, no. 3 (2013): 623-628.

31. Alexander, “Causing the Conditions,” 623.

32. Ibid, 626.

33. Ibid, 626-7.

34. Ibid, 628.

35. Of course, were there international legal institutions that could hear a case against such a state, that would be preferable.

36. Jeff McMahan, “Who Is Morally Liable to Be Killed in War,” *Analysis* 71,

no. 3 (2011): 555-6. McMahan uses the example of the soldier who is armed and dressed as a combatant but has no intention to fire a weapon and is hoping to escape the battle. This individual does not objectively pose an unjust threat of harm to others but is morally responsible for appearing as such.

37. Ironically, some proportionate responses from the security forces might also be used to negatively affect the public perception of the occupying regime. A practical dilemma is possible where both proportionate and disproportionate responses will lead to equally negative publicity.