Police Obligations to Aggressors with Mental Illness

Abstract: Police killings of individuals with mental illness have prompted calls for greater funding of mental health services to shift responsibilities away from the police. Such investments can reduce police interactions with vulnerable populations but are unlikely to eliminate them entirely, particularly in cases where individuals with mental illness have a weapon or are otherwise dangerous. It remains a pressing question, then, how police should respond to these and other vulnerable aggressors with diminished culpability (VADCs). This article considers and ultimately rejects three potential approaches from the ethics of defensive force literature. It looks to improve on them by developing what I call the fusion account, which explains how vulnerability and diminished culpability fit together to provide moral grounds for extra protections from deadly force. The article’s final sections explore the policy implications of the fusion account for police administrators, officers, and the law.

Keywords: culpability, deadly force, distributive justice, police, mental illness, vulnerability

Author: Ben Jones, Assistant Professor of Public Policy and Research Associate in the Rock Ethics Institute (btj7@psu.edu)

***Please cite the final version in the Journal of Politics at https://doi.org/10.1086/726972***
In October 2020, protests broke out in Philadelphia, Pennsylvania, after the fatal police shooting of Walter Wallace Jr. His family called 911 for help because he was experiencing a mental health crisis. Video captured Wallace in the street pointing a knife at officers, who shot him after he failed to drop it (Dale 2020). Sent to help a vulnerable individual, police killed him instead.

Such killings by police are all too common. Each year police in the United States fatally shoot around 1,000 people, with more than one in five of them experiencing a mental health crisis (Washington Post 2023). The ethics of defensive force literature offers little direct discussion of police obligations in these cases. One question it takes up is that of killing innocent aggressors—those who pose an unjust intentional threat to life but lack moral responsibility and culpability for their threat (Otsuka 1994, 74; Quong 2009, 507). Thought experiments illustrating these encounters often have an artificial quality, like someone developing the intent to kill after being drugged against their will (Thomson 1991, 284). Though not impossible, such incidents are incredibly rare. The focus on them can distract from more likely scenarios involving aggressors who, even if not fully innocent, have severely diminished culpability due to factors like mental illness. Since encounters with these vulnerable populations occur regularly in police work, determining when deadly force is justified in such cases is an urgent normative task.

As evident from mass protests in Philadelphia and elsewhere, police killings of vulnerable individuals strike many as a moral failure, regardless of whether deadly force is justified under current law. This article examines that intuition, with a focus on aggressors with mental illness who pose an intentional threat of grave harm, like death or serious bodily injury. Drawing on the ethics of defensive force literature, it considers four ways to understand police obligations to these and other vulnerable aggressors with diminished culpability (VADCs):

(1) **Traditional account:** VADCs do not enjoy extra protections from deadly force.

---

1 Such aggressors act with intent to harm but are not culpable for that intention so lack mens rea (Rodin 2002, 92).
(2) **Culpability account:** innocent VADCs are not liable to deadly force.

(3) **Vulnerability account:** VADCs’ vulnerability renders them less liable to deadly force.

(4) **Fusion account:** VADCs enjoy extra protections from deadly force due to their vulnerability and diminished culpability.

On prioritarian and egalitarian grounds, I argue that the traditional account’s failure to provide VADCs extra protections gives us reason to question it. With their approaches to liability, the culpability and vulnerability accounts represent potential solutions, but both have shortcomings. As an alternative, I develop the fusion account, which pulls from aspects of the culpability and vulnerability accounts to offer a firmer basis for extra protections that apply to VADCs.

At the start, it is important to situate my argument within current debates over policing. Protests in response to high-profile killings by police in 2020 led to calls to reimagine public safety and limit police interactions, especially with vulnerable populations (Jones and Mendieta 2021, 2). Proposals include more robust services for vulnerable populations and teams of mental health professionals better suited than police to respond to individuals in crisis. Motivating these proposals is the idea that avoidable police violence stems from a broad range of societal failures. Indeed, in the US and elsewhere, policing often proves challenging because it occurs against a backdrop of inadequate funding and services for vulnerable populations (Morabito 2014).

Though actors besides the police have obligations to vulnerable populations, it remains important to understand police obligations to these populations. Even if more robust social services reduce the need for police interactions with vulnerable populations, occasions will arise (albeit less frequently) where a vulnerable individual has a weapon or is otherwise dangerous and police must respond. For instance, the CAHOOTS program in Eugene, Oregon—a service offering civilian crisis intervention teams often championed as an alternative to the police

---

2 CAHOOTS stands for Crisis Assistance Helping Out On The Streets.
(McLeod 2019, 1630)—reduces police contacts yet still relies on officers as the primary responders for individuals in crisis who are violent or have a weapon (Beck, Reuland, and Pope 2020). Unfortunately, perfect compliance and the elimination of violence only exist in ideal theory, not any actual society—not even those most proactive in addressing the social determinants of violence. So under favorable conditions, police still must be prepared to respond to VADCs. And despite growing calls to address the needs of vulnerable populations, those needs often remain unmet and many officers face the challenge of how to respond to VADCs under nonideal conditions. Police continue to have obligations to VADCs in such circumstances.

This article explores those obligations. It first examines two key features of VADCs—vulnerability and diminished culpability. The next four sections outline competing approaches to morally evaluating deadly force against VADCs, ultimately defending the fusion account. The following section addresses objections to the fusion account with a revised view of necessity. Prior to the conclusion, the final three sections consider the fusion account’s policy implications for police administrators, officers, and the law. Social science research and actual incidents of police violence against aggressors with mental illness inform the article throughout, with the goal of keeping the analysis tied to the ethical and policy challenges that such encounters pose.

Culpability and Vulnerability

Culpability and vulnerability are distinct but often linked concepts. Culpability refers to the extent that someone is morally blameworthy for wrongful or unjust action, meaning that it violates or attempts to violate other people’s rights. There are different degrees of culpability. Fully culpable aggressors lack moral justification, permission, or excuse for the intentional harm

---

3 Initial research on civilian crisis response programs has proved promising. A quasi-experimental study of the Support Team Assisted Response (STAR) program in Denver, Colorado, found that it led to crime reductions while decreasing contact with police and the criminal justice system for individuals in crisis (Dee and Pyne 2022).
they threaten or inflict on others (McMahan 2009, 159). For other aggressors, features like youth, mental illness, or intellectual disability diminish their culpability by limiting their deliberative agency and control over their actions (Rodin 2011, 83). Some exculpatory factors like serious mental illness can diminish an aggressor’s agency to such an extent that they lack any culpability for their threatened or actual harm. That limit case is an innocent aggressor.

As understood here, vulnerability refers to susceptibility to harm (Goodin 1985, 110; Mackenzie, Rogers, and Dodds 2013, 4). All humans are vulnerable—to disease, injury, aging, and other factors (Fineman 2008). But some are more vulnerable than others. Though certain physical and psychiatric conditions come with inherent vulnerabilities, social conditions are a significant driver of vulnerability (Scully 2013, 207–9). Societal structures often create and exacerbate the vulnerabilities of marginalized groups by failing to be responsive to their needs.

A feature can impact vulnerability but not culpability, and vice-versa. For instance, being Black in America where systemic racism persists comes with greater vulnerability to various harms, including police violence. But someone is not more or less culpable because of their race (though it can impact perceptions of culpability when bias is present).

Mental illness can diminish someone’s culpability while also making them more vulnerable given background social conditions. There is abundant evidence that the criminal justice system exacerbates vulnerability for this population. In the US, those with serious mental illness end up in prison and solitary confinement at higher rates than the rest of the population (Maruschak, Bronson, and Alper 2021; Simes, Western, and Lee 2022). Individuals with serious mental illness also are more likely to be killed by police—seven times more likely according to US data from 2015 (Saleh et al. 2018, 114).

---

4 Permissions apply to action not wrong to perform, justifications to permissible action for which there is also positive moral reason to perform it, and excuses to wrongful action where mitigating circumstances lessen or eliminate moral blame for performing it (McMahan 2009, 110).
So when an aggressor has mental illness, they possess a feature that both increases their vulnerability and diminishes their culpability. I define this class of aggressors as follows:

*Vulnerable aggressors with diminished culpability* (VADCs): aggressors with a feature (or features) beyond their control—like youth, mental illness, or intellectual disability—that significantly diminishes their culpability for their unjust threat and significantly increases their vulnerability to harm, including force, which makes them worse off.

There are a few points to note about this definition. First, it characterizes VADCs not as inherently worse off but as worse off in the social conditions in which they live—conditions marked by vulnerability.\(^5\) Second, due to features beyond their control, VADCs live under prolonged vulnerability to a range of harms including violence. Wholly culpable aggressors may face a greater risk of force in the moment they attack others, but this risk is fleeting and one they bring on themselves, so they do not qualify as vulnerable in the sense that VADCs are. Third, VADCs range from the partially culpable to the wholly innocent (innocent aggressors are a subset of VADCs with fully diminished culpability). Inevitably, there are marginal cases where it is debatable whether an aggressor’s condition *significantly* impacts their culpability and vulnerability. We can lack clarity on how to handle marginal cases (e.g., mild mental illness), while still arriving at firmer conclusions for more clear-cut cases (e.g., serious mental illness).

This article takes that approach, with the goal of offering guidance for an important range of cases involving aggressors with mental illness, even if tough cases remain.

**Proposal 1: Traditional Account**

As the prevailing approach in law, the traditional account has significant influence in governing deadly force. This view rejects that VADCs should receive extra protections from deadly force.\(^6\)

---

\(^5\) For more on the relationship between mental illness and individual welfare, see Carter (2023).

\(^6\) A less common view in law and philosophy is that VADCs with severe and permanent intellectual disability or mental illness should receive diminished protections from deadly force because their lives have less value than
Its position follows from excluding culpability and vulnerability in determining whether deadly force is justified. According to the traditional account, deadly force is justified if one reasonably believes that such force is necessary to stop an aggressor’s unjust threat of grave harm. As long as those conditions are met, deadly force is justified regardless of whether the aggressor is innocent, partially culpable, fully culpable, or vulnerable. Conditions such as insanity or mental illness ultimately do not impact an aggressor’s liability to defensive force (Fletcher 1973; Thomson 1991)—that is, whether they have forfeited their moral right not to be attacked and harmed by others (McMahan 2009, 10). Like anyone else, VADCs who pose a grave threat put themselves at serious risk of being liable to deadly force. The Model Penal Code adopts this approach (Dubber 2015, 157), which shows up in criminal law and provides legal justification for many police killings of aggressors with mental illness.

On its face, the traditional account has certain advantages. Sometimes it is difficult to determine an aggressor’s culpability and vulnerability, especially when police respond quickly to a threat. By treating aggressors the same regardless of their culpability or vulnerability, the traditional account avoids imposing on police and others unreasonable epistemic burdens (Klinger 2021, 125–26). Furthermore, though many recognize a role for culpability in determining an aggressor’s punishment after an attack, it is more controversial to claim that an aggressor’s culpability (or vulnerability) impacts their liability to defensive force to prevent an attack. An innocent aggressor still can violate others’ rights against being harmed. The traditional account makes the plausible, though debated, claim that one is justified in using proportionate and necessary force to fend off an innocent aggressor’s unjust attack.

“normal” human beings (Bouzat 1963, 272; Otsuka 1994, 92–93). Because this approach dehumanizes certain individuals and receives little consideration as a potential policy option, I do not consider it here. For critiques of that approach, see Fletcher (1973, 374–75) and Kaufman (2010, 87–89).
Despite its apparent advantages, the traditional account has shortcomings. Notably, some of the worries motivating it are exaggerated. It would be unfair to expect police to respond differently to VADCs who, based on an officer’s reasonable belief, do not appear vulnerable or to have diminished culpability. Yet in other cases—like when serving a mental health warrant or responding to a mental health crisis call—officers know that an aggressor is likely a VADC. So the idea that police have special obligations to VADCs does not entail unrealistic demands. We can rely on a standard like reasonable belief to determine when such obligations apply.

The more fundamental problem with the traditional account is that its one-size-fits-all approach leaves little room for a broadly shared intuition: it is morally preferable in the distribution of benefits to give priority to those who are worse off. This principle from prioritarian and egalitarian thought proves most compelling when there is a close link between the disadvantage in question and the benefit being distributed. Not all disadvantages are morally relevant for distributing a benefit like protections from police violence. It is doubtful that greater vulnerability to, say, wildfires provides moral grounds for extra protections from police violence. But for persons with mental illness, the harms to which they are more susceptible include police violence. Their vulnerability renders them worse off, suggesting that they should receive some priority in protections from police violence. That claim does not mean vulnerable lives always take precedence over others since many factors matter when morally evaluating force. Rather, more modestly, vulnerable lives should be prioritized all else being equal. The traditional account’s failure to accommodate that intuition gives us reason to explore if other accounts can.

Prioritarians give priority in the distribution of benefits to those absolutely worse off, whereas egalitarians give priority to those comparatively worse off (Parfit 2002). Though these positions can diverge, they point to similar conclusions regarding police obligations to VADCs, who are worse off both in absolute and comparative terms.
Proposal 2: Culpability Account

With its approach to liability, the culpability account offers a potential remedy to the traditional account. Developed and defended by Kimberly Kessler Ferzan (2005, 2012) and others (e.g., Alexander 1987), it treats culpability as a necessary condition for liability. The culpability account emphasizes that there should be a high bar for losing one’s right against being attacked. And culpability—engaging in morally blameworthy action—is a higher bar than what the traditional account proposes. Aggressors who lack culpability and are innocent still can be liable to deadly force according to the traditional account. In contrast, the culpability account understands innocent aggressors as not liable to deadly force, even if it is necessary to prevent them from killing innocent victims. This categorical rejection of the idea that innocent aggressors can be liable to deadly force does not suggest a complete prohibition on such force against them. For instance, if deadly force is necessary to prevent an innocent aggressor from killing several innocent victims, there can be a lesser evil justification to kill the innocent aggressor to prevent the greater harm of multiple innocent people being killed (Alexander 2005, 612).

Still, with that caveat, the culpability account grants innocent aggressors greater moral protections than the traditional account does, which has implications for VADCs—specifically, those who are innocent aggressors. Treating them as not liable to deadly force gives priority to the lives of certain VADCs, who by definition are worse off. So compared to the traditional account, the culpability account better accommodates the intuition that those worse off should be prioritized in the distribution of benefits, like protections from deadly force.

On closer inspection, though, the culpability account makes only limited room for that intuition, which casts doubt on whether it overcomes the traditional account’s deficiencies. The culpability account’s emphasis on innocence turns out to be a double-edged sword. It provides
moral protections to the wholly innocent but fails to extend these protections to aggressors—including those with diminished but partial culpability—who fall anywhere below that bar.

This point is evident in the claim by Larry Alexander (2013, 169), a defender of the culpability account, that liability to deadly force applies to “all culpable aggressors, no matter how low their level of culpability, if they threaten an act that could cause great harm.” Relatedly, Ferzan (2005, 744–45) stresses that an aggressor’s probability of killing someone can be quite low yet they remain liable to deadly force if they are culpable for intending such harm. So if an aggressor is at all culpable for a threat posing any risk of grave harm, defenders of the culpability account understand its protections against deadly force as no longer applying to the aggressor.

It is doubtful that most aggressors with mental illness qualify as innocent aggressors. Mental illness often diminishes VADCs’ culpability without eliminating it (Pickard 2015). Such VADCs occupy a worse-off position in society without the benefit of extra protections from deadly force, according to the culpability account. This view offers little to prioritize their lives.

Notably, some defenders of moral protections for innocent aggressors are reluctant to extend those protections to individuals with mental illness. When arguing that innocent aggressors are not liable to deadly force, David Rodin carves out significant exceptions in cases of mental illness. Only innocent aggressors whose mental “aberration … has its source entirely outside of the mental and psychological world of the subject”—like a blow to the head—are not liable to deadly force. But aggressors whose threats stem from factors internal to their psychology do not enjoy the same moral protections according to Rodin (2002, 94–95). In some cases, mental illness emerges from internal factors beyond individuals’ control and is so severe that it renders them not culpable for their aggression. These aggressors certainly appear innocent, yet for Rodin they exist outside the moral protections purportedly for innocent aggressors.
Even for innocent aggressors covered by moral protections outlined in the culpability account, its implications for law and policy are somewhat limited. Most defenders of the culpability account treat deadly force against an innocent aggressor as excused if such force is necessary to stop their grave threat (Ferzan 2005, 748). An excuse suggests that one should be exempt from blame and penalties. So though the culpability account sees innocent aggressors as not liable to deadly force, it is a mistake to conclude that this view calls for sanctions for officers who use deadly force against such aggressors when needed to stop their grave threat.

None of the points raised so far deny the culpability account’s central claim that innocent aggressors are not liable to deadly force. How to respond to innocent aggressors threatening grave harm, where deadly force is necessary to stop it, is a thorny question that involves competing moral considerations. Since causing harm is seen as more difficult to justify than allowing it, that consideration counts against using deadly force (Woollard 2015). At the same time, many believe that giving some preference to preserving one’s own life is permitted or justified (Quong 2020, 58–96). This article remains agnostic whether someone with only two options—kill an innocent aggressor or be killed by them—is morally excused, permitted, or justified in using deadly force. Importantly, all three positions suggest that moral blame and legal sanctions are inappropriate for someone who employs deadly force in such a dilemma.

Though the culpability account provides greater moral protections to VADCs than the traditional account, these protections prove limited and fail to cover most VADCs. As typically formulated, then, the culpability account runs into problems similar to those of the traditional account. It offers little to prioritize the lives of many VADCs who, it seems, should receive some priority due to their vulnerability and worse-off position in society.

---

8 Alexander (1987, 1187–88) suggests that the excuse to kill an innocent aggressor applies to those acting in self-defense but not third parties. It is less clear that Alexander would extend this principle to police given their duties to prevent harm (Fabre 2007; del Pozo 2023), a point that he seems to implicitly recognize (Alexander 1987, 1183).
Proposal 3: Vulnerability Account

The third proposal comes from Jeff McMahan. In the ethics of defensive force literature, McMahan (2005, 2009) is best known for the moral responsibility account, which argues that someone’s liability to defensive harm varies directly with their moral responsibility. Moral responsibility encompasses culpability but is a broader concept since one can be morally responsible, without being culpable, for permissible action that poses a risk of foreseeable harm, like driving a car (McMahan 2005, 394–95). McMahan (2009, 198–202) also suggests a complementary yet distinct proposal concerning liability that has received less attention: aggressors’ vulnerability renders them less liable to deadly force (see also Fabre 2018). This section focuses on that proposal, which I call the vulnerability account, because of its implications for VADCs. The vulnerability account singles out greater susceptibility to harm as reason to grant VADCs extra protections from deadly force.

This idea comes up in McMahan’s discussion of child soldiers from *Killing in War*:

> [W]hen just combatants could use lesser force against child soldiers without seriously compromising their ability to achieve their just aims, they may be morally required to fight with restraint, even at greater risk to themselves…. I suspect that any commander would earn the respect of his troops if he were to order them to take additional risks to try to drive back, incapacitate, subdue, or capture child soldiers, while sparing their lives. (McMahan 2009, 201–2; see also McMahan 2010)

For McMahan, just combatants have an obligation to make extra efforts to avoid deadly force against child soldiers, even if doing so puts combatants at greater risk. This claim has intuitive appeal. Most would be troubled if combatants treated child soldiers like any other adversary.

Child soldiers make moral demands on combatants that other aggressors do not. In explaining why, McMahan notes that their diminished culpability and responsibility count in

---

9 Recent book-length studies on the ethics of defensive force by Frowe (2014) and Quong (2020) never mention vulnerability. Lazar (2015, 101–22) does see vulnerability as relevant for evaluating force in war but only considers it for noncombatants, not aggressors. Bolinger (2021, 143) also discusses vulnerability but understands it differently from how it is used here. For her, vulnerability means the absence of rights against being harmed.
favor of restraint, but he especially emphasizes their vulnerability: “[T]hese soldiers are children—that is, individuals who have hardly had a chance at life and have already been terribly victimized…. [W]hen child soldiers are conspicuously young, there is moral reason to exercise restraint simply because of their special vulnerability to exploitation and loss” (McMahan 2009, 201–2). That description applies to children in war zones vulnerable to being coerced into military service and suffering grave harm. Given their vulnerability, diminished culpability, and worse-off position due to societal failures to protect them, child soldiers clearly are VADCs.

On its face, child soldiers represent a compelling case in support of the vulnerability account. This example, though, is not the cleanest test of the vulnerability account. As VADCs, child soldiers are characterized by vulnerability and diminished culpability, which both could impact liability. It is important to disentangle the respective impact of these factors.

Cases of aggressors characterized by vulnerability or diminished culpability, but not both, help with that task. Consider first a young child who finds a deadly weapon and wields it in a way that threatens others, while in a society with robust measures to protect children so that overall they are among those least vulnerable to force and other harms. The child is not vulnerable like a VADC, yet their youth leaves them barely culpable. The vulnerability account suggests that an innocent potential victim is justified in using less restraint against this child than against a vulnerable child aggressor, like a child soldier, all else being equal. But that claim is doubtful. Whether the aggressor comes from vulnerable conditions appears to have little impact on liability in these unfortunate circumstances. Their status as a young child, with significantly diminished culpability, demands similar restraint in both cases.

Now consider a fully culpable adult aggressor who is a member of a vulnerable group. As child soldiers illustrate, conditions of vulnerability often come with abuse, trauma, and duress,
which all can diminish culpability. But it is possible to not experience factors diminishing culpability while being part of a group that overall is more susceptible to harm. The vulnerability account treats a culpable and vulnerable aggressor as less liable to deadly force than an equally culpable but not vulnerable aggressor. So an innocent potential victim is obligated to exercise greater restraint and accept greater risks against the former than the latter. That claim is at odds with a micro-level perspective focused on justice in individual interactions, which prioritizes protecting the innocent over the culpable who are at fault for the increased risk of harm. The micro-level perspective is incomplete, though. It leaves out the context for the encounter: systemic injustice that makes certain groups more susceptible to harm and worse off. From a macro-level perspective focused on distributive justice broadly in society, vulnerable groups—like the one the aggressor is a part of—should on prioritarian and egalitarian grounds receive remedies for the greater harms that they face. But even with those macro-level considerations, the vulnerability account’s prescriptions in this case remain questionable. After all, it is far from clear that the burden of providing remedies for systemic injustice should fall heavily on innocent persons threatened by vulnerable yet culpable aggressors. That responsibility does clearly lie with state institutions charged with advancing distributive justice, which should strive to fulfill their task in ways consistent with micro-level considerations of justice. The vulnerability account makes the controversial move of relying on macro-level considerations for a micro-level domain, liability, concerned with justice in individual encounters (see Burri 2022).

These examples cast some doubt on the vulnerability account. In cases disentangling vulnerability and culpability, the vulnerability account has less intuitive appeal than for child soldiers and other VADCs. Given these limitations, the next section looks to identify a firmer moral basis for extending to VADCs more extensive protections from deadly force.
Proposal 4: Fusion Account

Though distinct from the culpability and vulnerability accounts, the fusion account draws on both. Like the culpability account, it treats culpability as a key factor for determining aggressors’ liability to deadly force. Like the vulnerability account, it treats all VADCs—not just the fully innocent—as enjoying extra protections from deadly force. Its approach to liability supports that conclusion, but the fusion account also cautions against relying too heavily on this concept. In the ethics of defensive force literature, there is often a hyperfocus on liability in hypothetical scenarios disconnected from background social conditions and institutions. Protections from deadly force go beyond the moral protections bound up in the concept of liability, and the fusion account emphasizes institutions’ responsibilities to foster such protections.

Let’s begin with the fusion account’s approach to liability. In the child aggressor example from the previous section, the aggressor’s diminished culpability provides reason for restraint regardless of their vulnerability. This example suggests that diminished culpability reduces liability to defensive force, which is how the fusion account understands the relationship between liability and culpability. Its approach departs from the standard culpability account already considered, which adopts a threshold view holding that any low level of culpability makes one just as liable to deadly force as a fully culpable aggressor. That view of liability has the effect of excluding many VADCs from its moral protections. In cases of child soldiers or aggressors with mental illness where additional restraint seems appropriate, even if the aggressor bears a low level of culpability, the fusion account has the advantage of matching that intuition. It understands these aggressors as less liable to deadly force due to their diminished culpability.\footnote{Quong (2012, 50) calls this the “variable culpability account.” The next section addresses his objections to it.} \footnote{The fusion account parallels the moral responsibility account mentioned in the previous section. In the former, liability varies with culpability; in the latter, liability varies with moral responsibility. Since moral responsibility is a broader concept that encompasses culpability, I understand the fusion account’s approach to liability as not in}
The fusion account’s view of liability has the further advantage of advancing macro-level distributive justice goals without conflicting with notions of justice at the micro-level—a problem with the vulnerability account. Because society fails to provide support that vulnerable groups like those with mental illness need, they are more likely to find themselves in situations involving force that could have been avoided. VADCs face a greater risk of harm than others who are *equally culpable* but do not live in the same conditions of vulnerability. This dynamic is similar to Black and White Americans using drugs at comparable rates but the former ending up in prison for it at much higher rates. For VADCs, low culpability carries the heightened risk of a deadly encounter. By providing greater protections to those with diminished culpability, the fusion account has the indirect effect of disproportionately benefiting vulnerable groups.

Some question the relevance of distributive justice considerations for liability and the ethics of defensive force generally (Quong 2020; Burri 2020, 2022). The fusion account shows that it is possible to have an approach to liability that advances broader distributive justice goals while focusing on a consideration, culpability, apt for evaluating justice at a local level. In sum, a radically novel approach to liability is not required to broaden protections for VADCs.

It is at the level of institutions that the distributive justice principle of prioritizing the vulnerable offers more direct guidance. In McMahan’s discussion of child soldiers—the source for the vulnerability account—he emphasizes that combatants should take on extra risks to protect these children (McMahan 2009, 198–202). That point is correct but incomplete. Those who take on extra risks to protect VADCs act admirably, yet it would be morally preferable to have in place measures that lower the risks for all parties. On prioritarian and egalitarian

---

conflict with the moral responsibility account, while remaining agnostic on perhaps the moral responsibility account’s most controversial claim—one is liable for permissible action whose small risk of harm eventuates (e.g., driving a car responsibly). I take no position on liability in such cases, which fall outside the scope of this article.
grounds, the fusion account recognizes that institutions have obligations to take steps to reduce the risk of grave harm for vulnerable groups before encounters even occur.

The police are among the institutions with such obligations. The literature on distributive justice often overlooks the police, yet their activities have clear consequences for the distribution of harms and benefits in society (see Monaghan 2021, 2023). Though the proper role of police is contested, one of their most widely recognized functions is to promote peace or public safety (Kleinig 1996, 22–29; Hunt 2019, 24–26). Law, policy, and tactics impact how that public good is allocated. To ensure its fair distribution, principles of distributive justice must inform policing.

Often various institutions bear responsibility for a group’s vulnerability to harm, even in a discrete area like deadly force. Since many factors contribute to conditions of vulnerability, police agencies by themselves cannot remedy the problem. But as an institution whose current practices often exacerbate vulnerability, police have a role to play in addressing it.

Police agencies are responsible for ongoing study of how best to prioritize the vulnerable in policy and practice. Such study should inform the development of training and tools to help officers uphold that principle in the field. In departments that fulfill these responsibilities, police dedicate significant time to training and preparation so that their responses to threats are just and effective. And the training officers receive in use of force and de-escalation comes with higher expectations. Compared with civilians, officers have more extensive and stringent obligations to protect the lives of VADCs.

Though this article focuses on aggressors with mental illness, it is worth noting that the principle of prioritizing the vulnerable also has implications for debates over racial disparities in police violence. Black Americans live under conditions of vulnerability where they are more likely to be harmed by police (Mummolo 2018; Edwards, Lee, and Esposito 2019; Nix and
Shjarback 2021). A commitment to prioritizing the vulnerable requires police and democratic institutions to pursue efforts to protect Black lives, which could include steps to reduce bias among officers, end the overpolicing of Black communities, and address the social determinants of violence through community investment (see Shelby 2016; Butler 2017). The fusion account proves consistent with the idea voiced by Black Lives Matter that, given deep inequities produced by persistent injustice, differential treatment is sometimes needed to advance justice.

**Objections and a View of Necessity to Address Them**

By treating liability as varying with culpability, the fusion account faces two related objections: its protections to VADCs are meaningless in practice or lead to implausible conclusions. An example illustrates these worries. Imagine that a barely culpable VADC attacks an innocent victim, who will be killed for certain unless they kill the VADC. In this case, it is impossible to divide harm between the victim and the aggressor. If the fusion account treats deadly force by the victim as justified on the ground that indivisible harm should fall on the more culpable party, the VADC does not appear any less liable than a fully culpable aggressor. Both are liable to deadly force (Quong 2020, 103–4). Conversely, if the fusion account avoids that conclusion by treating the VADC as liable to a lesser amount of defensive force—say only force that would cause moderate injuries—the victim no longer is justified in using the deadly force needed to stop the threat. On that interpretation, the fusion account requires the victim to allow the unjust aggressor to kill them, which strikes many as an implausible conclusion (Quong 2012, 50–51).

Such a demand would be unreasonable, and it is a mistake to attribute that view to the fusion account. But the alternative interpretation—the VADC in the example is liable to deadly force (Quong 2020, 103–4). Conversely, if the fusion account avoids that conclusion by treating the VADC as liable to a lesser amount of defensive force—say only force that would cause moderate injuries—the victim no longer is justified in using the deadly force needed to stop the threat. On that interpretation, the fusion account requires the victim to allow the unjust aggressor to kill them, which strikes many as an implausible conclusion (Quong 2012, 50–51).

---

12 Quong raises this objection against the view that liability varies with moral responsibility (the moral responsibility account), but his objection also applies to the view that liability varies with culpability (the fusion account).
force—does not mean that their liability matches that of a fully culpable aggressor. That objection assumes that liability varies in only one sense: the magnitude of force that justifiably can be imposed on someone. But there is another sense of being more or less liable to defensive force: the range of circumstances in which one is liable to a particular level of force. In the above example, it is certain what will stop the aggressor’s threat, but many and likely most cases of aggression lack the luxury of such certainty. Often there is ambiguity over the level of force needed to stop a threat. It is these cases, I argue, that VADCs’ diminished culpability becomes most relevant in reducing their liability and the level of force justified against them. A revised view of necessity helps explain why and complements the fusion account’s approach to liability.

Both necessity and proportionality are widely recognized conditions that must be met to justify force. Proportionality prohibits force whose harm greatly exceeds the unjust threat that it aims to prevent, and necessity prohibits force that, even if proportionate, causes more harm than required to stop the threat. In cases of uncertainty, the standard view of necessity prohibits force that would cause more harm than other options with an equal or better chance of stopping a threat, based on one’s reasonable belief. This rule permits questionable uses of force, however. A proportionate option with a better chance of stopping a threat than all alternatives is justified even if another option with nearly the same chance of success would inflict much less harm. So if deadly force has a 95% chance of stopping a VADC’s grave threat and de-escalation tactics a 94% chance of success, police are justified in using deadly force according to the standard view of necessity. Yet that seems wrong—police should take on a small additional risk to avoid the grave harm that deadly force would inflict on the VADC (Lazar 2012, 12; McMahan 2016, 187).

An alternative developed by Lazar (2012, 2018) offers more plausible guidance in such cases (see also McMahan 2016). It understands necessity as only justifying force that minimizes
“morally weighted harm” compared with other options (Lazar 2012, 6–14). Morally weighted harm accounts for the range of ethical considerations that matter for evaluating force under uncertainty. For instance, a greater likelihood of preventing unjust harm would lower an option’s morally weighted harm. And causing grave harm to an aggressor with diminished rather than full culpability would increase an option’s morally weighted harm. Such considerations can pull in opposite directions. The concept of morally weighted harm captures those tradeoffs. That point is important for VADCs since allowing their unjust threats poses obvious harm, yet their diminished culpability increases the morally weighted harm of using force against them.

To illustrate, consider an aggressor with mental illness who pulls a gun on a nearby officer. Besides deadly force, no option is likely to stop this grave threat. The officer could try tackling the aggressor, but that tactic has a low chance of success. The significant gap between lethal and less lethal options in their expected effectiveness strengthens the moral case for deadly force. Avoiding deadly force carries the high likelihood of grave harm to an innocent party, whereas using it carries the high likelihood of grave harm to an aggressor with diminished culpability. Though both harms are significant, it is unlikely that that the latter exceeds the former from the perspective of morally weighted harm. Deadly force by the officer therefore would qualify as necessary and justified.

Now consider the example to open this article, the police killing of Walter Wallace Jr. In contrast to the previous example, Wallace had a knife. Despite qualifying as a deadly weapon, a knife is significantly less deadly than a firearm. In fact, for the last decade of available data (2012–2021), only four officers in the US were killed in knife attacks according to the Federal Bureau of Investigation (2022). That low number cannot be explained by US officers’ being so quick to shoot aggressors with knives. Officers in the United Kingdom also rarely die in knife
attacks but are trained to respond without relying on a firearm (Zimring 2017, 100–2). Such data suggest that police resorted to deadly force against Wallace when less lethal tactics, even if not guaranteed to succeed, had a real chance of success. That consideration and Wallace’s diminished culpability increase the morally weighted harm of deadly force so that it likely exceeds that of other options, suggesting that such force would violate necessity.

Police killings of VADCs with knives often are treated as legally justified in the US (e.g., Danahy 2019), a conclusion easier to arrive at with the standard view of necessity. Since a knife is a deadly weapon, deadly force is generally understood as a proportionate response to a knife attack. In addition to being proportional, deadly force against an aggressor like Wallace would satisfy the standard view of necessity as long as officers reasonably believed that such force offered any slight advantage in likelihood of stopping the threat over alternatives like de-escalation.13 That claim is not implausible given deadly force’s effectiveness at incapacitating.

The fusion account’s view of necessity offers a rationale for greater restraint. For aggressors generally, this view can require officers to pursue measures that involve taking on additional risk than that associated with using deadly force, especially if the extra risk is small. A significantly greater chance of stopping the aggressor’s threat and preserving their life morally outweighs a small additional risk to officers. Since killing aggressors with mental illness and other VADCs involves greater morally weighted harm, police in encounters with them can be obligated to take on greater risks than what necessity would demand in encounters with fully culpable aggressors, all else being equal. Correspondingly, VADCs are liable to deadly force in a narrower range of circumstances than fully culpable aggressors.

13 Relatedly, the culpability account would deem Wallace liable to deadly force as long as he bore any culpability for his threat of grave harm (even if low) and his threat had any chance of success (even if low). See Alexander (2013, 169) and Ferzan (2005, 744–45), both discussed above in Proposal 2: Culpability Account.
Whether there is an obligation to take on extra risk depends on the availability of options to meaningfully divide risk between the defender and aggressor. In the encounter with Wallace, de-escalation tactics with the option of less lethal force would not have eliminated all risk to the officers and VADC but would have distributed risk more justly than the actual response. Other scenarios preclude the division of risk because the nature of the conflict nearly guarantees either the defender or aggressor will be killed—avoiding deadly force means bearing all risk of harm. The fusion account rejects blame or sanctions for officers who use deadly force in such difficult circumstances. Its view of necessity sees aggressors’ culpability as one but not the only factor to consider when comparing the morally weighted harm of potential responses. The likely effectiveness of available options also impacts whether deadly force is morally justified.

**Implications for Police Administrators**

In many places, police interactions with persons with mental illness would look much different if the fusion account guided them. A case illustrating the status quo drives home that point: the 2020 fatal police shooting of Adrean Stephenson in Sarasota, Florida. Stephenson was a frail 63-year-old woman who lived with chronic pain. She had a history of mental illness and was suicidal the day that police killed her. Video of the incident shows two deputies outside with their weapons aimed at Stephenson, who after placing a fillet knife to her neck pointed it at police. The deputies commanded her to drop the knife and then one fired his Taser. When that failed, the other officer quickly resorted to shooting and killing Stephenson (Munoz 2020). Afterward, Sarasota County Sheriff Tom Knight called it a “tragic situation,” bemoaning the lack of mental health services. He emphasized that his deputies acted properly and only used deadly force after exhausting “all attempts at de-escalation” (*Venice Gondolier* 2020).
That response is not surprising since criticizing the officers who killed Stephenson would call into question the policies and training in place. Even when police administrators require officers to receive training in verbal de-escalation for encounters with individuals experiencing a mental health crisis, such training often remains paired with a version of the 21-foot rule. That rule instructs officers to use deadly force against aggressors with a weapon less lethal than a firearm, like a knife or club, who get within 21 feet (Stoughton, Noble, and Alpert 2020, 168–71; Jones 2022, 370). When officers are taught the 21-foot rule or, more generally, to resort to deadly force if verbal de-escalation fails and an aggressor still approaches, attempts at de-escalation often do not last long. So the deputies who killed Stephenson may have acted in accordance with their training and departmental policies. But the more important question is whether measures to ensure extra protections for the vulnerable were in place to begin with.

Police administrators can advance that goal through adopting policies, training, and equipment that raise the bar for when deadly force is necessary in encounters with VADCs. When police administrators take those steps, incidents like the one with Stephenson will likely end differently. Consider, for instance, how UK police responded in 2011 to a more dangerous aggressor with mental illness—a large man wielding a machete. In contrast to most US officers, UK officers receive more robust training in de-escalation tactics to resolve such incidents without deadly force. So when UK officers responded to a VADC with a machete, video shows their proactive efforts to maintain distance while keeping him surrounded until additional officers with shields arrived (CBC 2016). This equipment rendered deadly force unnecessary and allowed police to disarm the aggressor through other tactics. Wide-scale adoption of such tactics by US police would especially benefit those with mental illness, who are more likely to be armed with a knife when killed by police (Saleh et al. 2018, 114).
These reforms would save the lives of other aggressors with a weapon less lethal than a firearm. Indeed, extra efforts to protect VADCs can raise the bar for when deadly force is necessary in other contexts. That outcome mirrors how measures to create a more accessible environment for those with disability have broader benefits. For example, curb cuts make it easier for those in wheelchairs to navigate on and off sidewalks, while benefitting cyclists, too.

Though the fusion account has policy implications whose benefits would extend beyond VADCs, it is a mistake to see it as indistinguishable from just calling for greater restraint against all aggressors. The fusion account also emphasizes giving priority to VADCs when police administrators make decisions on how to deploy limited resources.

Another potential implication for policy illustrates that point. In addition to changes in training for officers generally, the fusion account may require police administrators to develop teams dedicated to further reducing the need for deadly force against VADCs. Through intensive training, specialized teams would have expertise in de-escalation with corresponding equipment to help resolve crises without deadly force. Such units likely would be a scarce resource that could not be deployed in response to every aggressor. Police administrators would have an obligation to prioritize protecting VADCs when using that resource.

Calling for investments in such a resource hardly seems unreasonable given current practices. Law enforcement in the US has made enormous investments in training teams to conduct militarized drug raids without the same urgency to reduce police killings of VADCs. This state of affairs is far from inevitable. In a world where police administrators engaged in proactive efforts to protect the most vulnerable, police units highly effective in resolving encounters with VADCs through nonlethal means would be the norm, not SWAT teams that often exacerbate the vulnerability of already vulnerable groups (see Mummolo 2018).
What efforts police administrators in the US have made to protect VADCs remain insufficient. The most common approach is Crisis Intervention Team (CIT) training, which a minority but growing number of agencies have adopted. CIT teaches officers about mental illness, de-escalation, and coordinating with mental health resources in their community (Rogers, McNiel, and Binder 2019). Though increased interest in CIT is a welcome step, it fails to go far enough, especially when outdated tactics like the 21-foot rule remain in place. Notably, the Sarasota County Sheriff’s Office (2018) boasted that all its “law enforcement members … are … trained in crisis intervention and de-escalation tactics.” The deputies who responded to Stephenson had a combined 120 hours of CIT training, yet that failed to prevent them from killing a 63-year-old woman in a mental health crisis (Henning 2020). Research finds little evidence that CIT training reduces the risk of injury or death for individuals with mental illness in police encounters (Rogers, McNiel, and Binder 2019; Fagan and Campbell 2020, 998–99; Marcus and Stergiopoulos 2022). These findings suggest that police administrators have an obligation to pursue strategies beyond just CIT training to protect aggressors with mental illness (see, e.g., Engel et al. 2022).

Implications for Officers

Due to failures of police administrators, some officers end up in situations where they lack the most effective training and tools available to protect the lives of aggressors with mental illness. But despite operating under less-than-ideal conditions, officers still have special obligations to VADCs. Deficient policies, training, and equipment mitigate but do not eliminate the blame falling to officers who fail to exercise greater restraint in encounters with VADCs. Even against the backdrop of institutional failures, the fusion account makes demands on officers.
It can require officers to take on extra risks to preserve the lives of VADCs. Consider the police encounter with Stephenson. Though the two deputies lacked training and equipment that would have been helpful in avoiding deadly force, they still could have done more to save Stephenson’s life. They resorted to deadly force quickly, spending less than a minute on other tactics (Munoz 2020). Given that Stephenson posed no immediate threat to anyone besides the officers, they could have prolonged attempts at de-escalation by repositioning and maintaining distance. Even if those tactics imposed a small additional risk on the officers, there is the moral expectation that they take that risk to protect life, especially for someone with mental illness.

Imagine the same incident, except the woman holding the knife was a close relative of the deputy who fired his gun. Would he have been so quick to shoot? Probably not. If faced with a loved one wielding a knife, most would be far more reluctant to use deadly force than the deputy who shot Stephenson. In fact, many would say that they are obligated to exercise greater restraint in this scenario. For at least some aggressors, then, we recognize our obligation to take on extra risks to save them. The fusion account applies a similar principle to police who face VADCs. That principle holds for Stephenson as well as for other VADCs more dangerous than her—like the mentally ill aggressor with a machete in the UK—where it is uncertain whether deadly force is needed to stop their threat.

What if an officer encounters a VADC threatening an innocent civilian? Though much depends on the details, generally there is a stronger intuition in favor of police deadly force when an innocent civilian is at risk. That intuition aligns with the fusion account’s view of necessity, which takes into account the significant harm of killing an innocent civilian. We should be wary, though, of jumping to the conclusion that any small but nontrivial risk to an innocent civilian by a VADC justifies deadly force. In the example that opened this article, the police shooting of
Walter Wallace Jr, his mother rushed into the street to try to stop her son from threatening police with a knife (Dale 2020). Some might justify police deadly force on the ground that, even if officers had an obligation to take on extra risks to protect an aggressor with mental illness like Wallace, shooting him was necessary to protect his mother. That justification is morally dubious, though. His mother willingly put herself in harm’s way to save her son, suggesting that she never would want that risk treated as a justification for killing her son. If someone willingly takes a risk to protect a VADC and that risk remains mostly limited to the person choosing it, officers should avoid treating it as a justification for deadly force.

The fusion account also implies obligations for officers in contexts beyond just encounters with VADCs. If a department has failed to implement policies, training, and equipment that would help save the lives of individuals with mental illness, its officers should lobby their superiors for such changes. Relatedly, they should seek out training grounded in evidence-based strategies for safely interacting with those experiencing a mental health crisis.

**Implications for the Law**

Police administrators and officers have obligations to make extra efforts to protect VADCs and, ideally, would willingly fulfill those obligations. But whether they do is a responsibility that, in the US and elsewhere, ultimately falls to democratic institutions that oversee the police. Democratic institutions have both the power and responsibility to implement laws and mechanisms of accountability for police that prioritize protecting vulnerable populations.

This issue came before the US Supreme Court in *City and County of San Francisco v. Sheehan* (2015). In 2008, San Francisco officers shot Teresa Sheehan multiple times in a group home for individuals with mental illness where she lived. Sheehan had threatened two officers
with a knife after they entered her private room where she was alone. The officers retreated but then shot her after taking the questionable step of reentering her room before the arrival of more officers, who could have helped avoid the need for deadly force. The Supreme Court granted the officers qualified immunity on the ground that they did not violate “clearly established” law. It also refused to rule on a question that lower courts had split on: Does the Americans with Disabilities Act’s (ADA) requirement that public entities provide reasonable accommodations to those with disability—including psychiatric disability (i.e., mental illness)—apply to police responding to individuals with mental illness who are armed and violent?

I am not concerned here with the interpretive question of how to read the ADA but the moral question of whether the law generally should require police to make accommodations to protect the lives of aggressors with mental illness. On the moral question, the fusion account makes clear that the answer is yes. Police have an obligation to take extra steps to protect the lives of VADCs. Given the state’s interest in protecting life, it has strong reason to put in place laws that promote this obligation and hold police accountable when they fail to fulfill it.

If such laws were in place, police could have faced legal penalties for the incidents discussed above—the shootings of Walter Wallace Jr, Adrean Stephenson, and Teresa Sheehan. Police were ill prepared to resolve these encounters through less lethal means, though training in those methods was available. The failure to implement such measures shows a troubling lack of concern for those with mental illness and other vulnerable individuals. Democratic institutions can and should push police to adopt policies, training, and equipment to address that failure, as well as provide the resources needed for their adoption.

The fusion account’s implications for the law intersect with its implications for police administrators and officers. Too often, police are a barrier to reform. But rather than stand in the
way of legislation to better protect VADCs, they should support it (see Monaghan 2017, 229–30). Police leaders who use their influence to advance such change, despite opposition from others in law enforcement, would demonstrate their commitment to protecting the vulnerable.

Conclusion

Police in the US kill individuals with mental illness at an alarming rate. Various factors are to blame, including inadequate mental health services that put police in difficult situations. Current police practices, though, exacerbate the problem. Many departments lack policies, training, and equipment that would reduce harm to those with mental illness and other vulnerable populations. For aggressors with mental illness, their vulnerability coexists with their diminished culpability, considerations that together call for more extensive protections from deadly force. The dearth of such protections in law and practice represents a moral failure, which police and democratic institutions must remedy if they are to fulfill their obligations to those with mental illness.

Acknowledgments

Earlier versions of this article were presented at the 2020 Rocky Mountain Ethics Congress, 2020 Association for Political Theory Conference, 2021 Contemporary Challenges for Just War Theory Conference at Temple University, and Penn State Bioethics Colloquium. I am grateful to those audiences for their suggestions, and particularly to Lee-Ann Chae, Elizabeth Lanphier, Désirée Lim, Michele Mekel, Jake Monaghan, Alex Tuckness, and Megan Wright. I also thank three anonymous reviewers for their exceptional feedback, which proved immensely helpful.
References


CBC. 2016. “UK Cops Disarm Man Wielding a Machete.”


*City and County of San Francisco v. Sheehan.* 2015. 575 US 600.


https://www.washingtonpost.com/graphics/investigations/police-shootings-database/

(accessed April 7, 2023).
