The Voluntary Act Requirement

By Gideon Yaffe

Introduction

It is abhorrent to punish someone for something he did not do. This feeling motivated Illinois Governor George Ryan in 2000 to commute the sentences of many defendants awaiting executions. The idea that the state should not just *punish* someone, but *execute* him, for something he did not do was more than Ryan could bear. It is rarely recognized, however, that there are two different ways to be punished for something one did not do. In the first kind of case, which is the sort that was on Ryan's mind, there is a doing, a genuine action, performed by *someone*—a murder, for instance—and worthy of punishment; the trouble is that the person being punished is not the very one who performed the awful act. In the second kind of case, there is no mistake of identity, but there is also no genuine action; the punished person is not punished for something he *did*, but, instead, for merely being a certain way—for "appearing" drunk on the highway after the police carried him bodily from his home to the highway before arresting him, as in Martin v. State (31 Ala. App. 334 (1944)), for instance. In such cases, someone is punished for something that cannot be attributed to his agency, something with respect to which he is passive. If it's wrong to punish in the first kind of case, isn't it just as wrong to punish in the second? In both, the punished person did not *do* anything prohibited.

Laws that enact the presumption of innocence are aimed at preventing conviction of people different from those who actually committed the relevant crime. Such laws enshrine our abhorrence of punishment of the innocent, the kind of

abhorrence that motivated Ryan. The criminal law's Voluntary Act Requirement (VAR) is often seen as enshrining in law the injustice of punishment in the second class of cases where there is no genuine action by the defendant on which to predicate criminal liability. The requirement is statutorily codified in many jurisdictions (Cf. Model Penal Code §2.01(1), critically discussed in Husak (2007)) and accepted as a foundational principle of law in jurisdictions in which it is not written into a statute. The VAR should be understood as follows: *A conviction of a defendant for crime C is justified only if (1) There is a voluntary act the performance of which is necessary for C's occurrence (given the statutory definition of C) and (2) The defendant has been shown (typically, beyond a reasonable doubt) to have performed such a voluntary act*. Under the VAR, we cannot criminalize in the first place, and so cannot legally punish, the meeting of a set of conditions that can be met without the performance of a voluntary act.

The VAR is a legal representation of a legitimate moral idea. There is something morally objectionable about violating it, as it would have been violated had the trial court's conviction of Martin for appearing drunk in public not been overturned. The question is what, exactly, the underlying moral idea is. As we will see, many of the valid moral principles that one might take the VAR to legally enshrine fail to capture important aspects of it. Further, it will be argued that, perhaps surprisingly, the moral idea that in fact justifies the VAR is the familiar one that what qualifies wrongful conduct for criminal punishment is the way in which that conduct manifests the defendant's objectionable mental states, such as bad intentions.

Section 1 explains some of the features of the VAR that need to be accounted for by a moral rationale for it. Section 2 discusses several attractive rationales and argues that each falls short of providing a fully satisfactory account. Section 3 offers an alternative, drawing on an appealing picture of the morally relevant relation between

criminal conduct and the mental states of the criminal, a relation which, it is argued, is at risk of being absent when the VAR is violated.

1. The Legal Doctrine

Although it is rarely made explicit, the law uses a definition of the term "voluntary act" that departs in significant respects from the ordinary concept. For legal purposes, a voluntary act is a *willed bodily movement*. To will, in the sense of relevance to the VAR, is not merely to mentally represent the bodily movement and for that bodily movement to be guided by that representation. If the mental state that guides the bodily movement is not "conscious"— a term explicitly used in the law—then that mental state is not a *willing,* or a *volition,* and the bodily movements it guides are therefore not voluntary actions. So the law provides two significant limitations in its definition of a voluntary act: only bodily movements count, and, of those, only bodily movements guided by *conscious* mental representations count. There are thus (at least) two kinds of voluntary action, in the ordinary sense, that do not count as voluntary actions in the law's sense: (1) voluntary thoughts (e.g. voluntarily thinking about a problem); and (2) bodily movements guided by unconscious mental representations of them.

The first of these two departures from the ordinary notion of a voluntary act is easily explained. There is something objectionable about criminalizing thoughts alone. Prohibitions on thoughts are intrusive violations of privacy, efforts at mind control, inconsistent with the goals and role of a liberal state. This idea gains legal expression in the idea that crimes consist of *both mens rea* elements *and actus reus* elements. The *mens rea* elements of the crime are mental states, thoughts, that the defendant must be shown

to have had for guilt; an intention to kill, for instance. But mature legal systems do not allow guilt merely on the strength of adequate proof of mental states. A*ctus reus* elements—further facts distinct from the mental states of the defendant—must be also be shown for guilt; that somebody died, and that the defendant caused that death, for instance. It is because the VAR is a restriction on the *actus reus*—among the *actus reus* elements must be a voluntary act—that the law defines the term "voluntary act" so as to exclude voluntary thoughts. The thoughts that are relevant to criminal liability are part of the crime's *mens rea* (and, in fact, none of them need to be shown to be voluntary). So, the fact that the VAR cannot be satisfied by proof that the defendant had a voluntary thought is merely an artifact of the formal structure of crimes—the division between *mens rea* and *actus reus*—that is, itself, a legal mechanism for barring crimes of pure thought.

The second departure from the ordinary notion of a voluntary act—the exclusion from the category of bodily movements guided by non-conscious mental representations of them—is more puzzling and in need of independent explanation. To see that this second category includes behaviors that we would ordinarily characterize as voluntary, consider the case of People v. Newton (87 Cal. Rptr. 394 (1970)). After a traffic stop and an altercation with police, Newton was shot in the gut. Immediately following, he shot and killed one of the police officers, and fled the scene, arriving shortly after at a hospital. Later he claimed not to remember the shooting, or traveling to the hospital, and claimed to have been unconscious for the crucial period of time beginning from the moment at which he was shot. Medical testimony supported the contention; a doctor testified that people suffering from traumatic injuries often engage in complex, motivated bodily movements in the absence of consciousness. Newton was convicted at trial, but his conviction was overturned on the grounds that the mental

states that guided his bodily movements—the finger movements on the trigger, for instance—were not conscious and so were not willings; but if they were not willings, then the relevant bodily movements were not voluntary, and so it would be a violation of the VAR to punish Newton for the officer's death. Newton's finger movements on the trigger were not likely to be purely reflexive; they were clearly goal-directed. Newton seems to have been *aiming* the gun at the officer, and so must have been mentally representing a particular goal, namely to shoot the officer, a mental representation that was involved in guiding his bodily movements. Many people would take such considerations to show that the relevant bodily movements were voluntary actions in the ordinary sense; but they were not voluntary in the law's sense.

Morally justifying the VAR requires explaining what justifies the exclusion of behaviors that we would ordinarily classify as voluntary acts. In addition, however, there are events that we would probably *not* classify as voluntary actions—in fact, they may not even meet *the law's* definition—but that the law treats as though they were. This is true, in particular, of omissions—failures to act—and of habitual actions. If a defendant is shown to have omitted a bodily movement (under certain conditions), or shown to have engaged in a habitual bodily movement, then that's good enough to comply with the restriction imposed by the VAR.

This is not as peculiar as it might seem. In many places in the law we find that X, which is different from Y, is treated for legal purposes as though it were Y. Under, for instance, the doctrine of Transferred Intent in homicide, a person who intends to kill one person and, acting on that intention, kills another instead, is treated *as though* he intended to kill the person he killed. The law does not make the mistake of holding that the defendant in such a case *really* intends to kill the person he kills; everyone recognizes that the death in such a case was, in a sense, an accident. Rather, it is that

the defendant is treated *as though* he had that intention since, it is thought, his actual intention was just as bad (or worse) than the one the law actually takes to be an element of an intentional homicide (namely, an intention to kill the person who was killed). Similarly, the law treats both omissions and habitual, often reflexive, acts *as though* they were voluntary acts for the purposes of the VAR, even though they may not be in fact.

Consider the case of omission first. There is an open question in the philosophy of action whether omissions are to be identified with willed bodily movements. Consider a mother who sits idly by while someone seriously injures her infant and makes no effort to stop the beating. Is her omission—the one we refer to when we say, "She failed to protect her own child"—to be identified with the bodily movement in which she *did* engage, the bodily movement, for instance, associated with sitting idly and watching? Perhaps. Or perhaps not; perhaps the action referred to by such a sentence consists only in *the absence* of the act of protecting her child, rather than that positive bodily movement that occurred in its stead. (For discussion and further references, see Bach (2010).) But whatever the outcome of this debate, a defendant in an omission case such as this need not *be shown beyond a reasonable doubt* to have willed the bodily movements in which she is engaging when she ought to have been protecting her child. And there's surely a good reason for this. Unless it speaks to the question of her capacity to protect her child, it would seem to miss the moral point of the VAR to acquit a defendant on the grounds that at the moment when she could have been protecting her child she was instead simply sitting idly, a bodily movement that she did not will. Precisely what seems morally salient is what *she did not do or will*, not whether or not she willed that which she was engaged in instead. The result is that the VAR is to be interpreted as allowing criminal liability in the absence of a voluntary action in the case of omissions meeting appropriate conditions. Like the exclusion by the VAR of

criminal liability in the cases bodily movements guided by unconscious mental states, the *in*clusion of criminal liability for the case of omissions must also be explained by an adequate account of the moral rationale of the requirement.

A word about habitual actions is also in order. Consider a defendant who has been trained by the military to spin around and fire immediately and without thinking on a threat behind him; this behavior has become, thanks to his training, habitual. Is he to be held guilty of a crime when, at the local firing range, he spins and fires on a person behind him who yells something threatening? Or would such a verdict violate the VAR? The bodily movements in cases such as this are routinely taken to provide an acceptable basis for criminal liability (cf. Model Penal Code §2.01(2)(d)). Noting that the defendant spun and fired only thanks to his military training will not serve to undermine the case against him through appeal to the VAR, although it might support a claim that he lacked some further mental state, such as an intention to kill, relevant to his degree of criminal liability. As suggested above, the law does not take this stance because it is believed that habitual actions are, in fact, *willed*. Nobody tasked with making legally binding decisions knows whether such actions are willed or not, or even whether some types are and some are not. It is an empirical question that would require careful investigation by psychologists, and neuroscientists. Normally, our ignorance about a feature of pertinence to criminal liability is enough to supply reasonable doubt, and thus enough to support an acquittal. But not in habitual action cases. If there is reasonable doubt about whether the defendant's bodily movement was voluntary *deriving from the fact that it is habitual* that reasonable doubt fails to undermine the case for guilt. We treat habitual bodily movements, that is, *as though* they were voluntary acts in the legal sense, even though we have no idea if they are in fact. This fact, also, about the VAR should be explained, ideally, by a moral justification of it.

To summarize, then, there are three features of the law's application of the VAR that should be accounted for by a justification of it. (1) There's no criminal liability for goal-directed bodily movements guided by non-conscious mental representations. (2) There is criminal liability for some omissions. And (3) there is criminal liability for habitual bodily movements even though there is reasonable doubt as to whether they are willed. An adequate rationale of the VAR will, first and foremost, explain why a voluntary action is necessary for justified criminal liability. But, ideally, the explanation will also entail (1), (2) and (3).

2. Insufficient Rationales for the VAR

Broadly speaking, there are two different kinds of moral rationale that can be given for a criminal law doctrine that draws distinction among defendants and then predicates differences in criminal punishment on the distinctions that it draws. Under the first kind of rationale, the doctrine is shown to draw a distinction on the basis of a property that makes a difference to desert of punishment. For instance, many offenses are graded by the mental states of the defendant. We punish more heavily someone who intentionally burns a neighbor's barn out of spite than someone who negligently burns it by forgetting to put out a campfire. This doctrine is explained by the fact that intentional wrongdoing is morally worse than negligent wrongdoing.

The second kind of rationale appeals to a moral norm governing the behavior of those who are to inflict punishment. Ordinarily, for instance, we think it wrong to punish someone who you lack good reason to think deserving of punishment *even if he is deserving of punishment*. To punish someone deserving without knowing for sure that he is deserving is not to give someone something other than what he deserves—if he's

deserving, he's deserving—but it is to nonetheless do something wrong; it is to violate a norm governing the conduct of punishers. This moral norm—do not punish unless *you know* the punished is deserving—explains, for instance, some criminal law doctrines pertaining to evidence: we often punish only one of two equally deserving defendants when one, but not the other, has been proven to be guilty by the admissible evidence.

Two of the possible rationales for the VAR discussed in this section—called here the "Evidentialist" and "Actual Invasion" rationales—are of the second sort: they identify moral norms governing punishers and claim that those norms would be violated were the state to fail to follow the dictates of the VAR. The other two—what will be called the "PAP" and "Reasons-Responsiveness" rationales—are of the first sort: they identify features on the basis of which the VAR distinguishes among defendants that speak to a difference in desert of punishment.

The Evidentialist Rationale

An adequate philosophy of mind will explain the fact that there is something private about our mental lives. We are wrong more than we like to admit about what's going on between our ears, but we nonetheless have access to it in a way that others do not. Despite their privacy, however, desert of punishment for wrongdoing turns crucially on the mental states of the wrongdoer. If we are to give punishment to only and all those who deserve it, we must somehow overcome the privacy of the mental in our assessments of who deserves to be punished. We have tools for doing this, perhaps most importantly the very same tools that we use to overcome the barrier of privacy in virtually every interaction that we have with any other human being: in the criminal law, that is, we exercise the very same tools for "reading minds" that we use when we

make eye contact with a stranger in the elevator. But when there is a great deal at stake in our judgments about what others are thinking—as there is when we inquire in order to decide whether and how to punish—we need a greater degree of certainty than we need in most human interactions. We need extremely good evidence of what is going on in the minds of those whom we might punish. And one might think that we necessarily lack that kind of evidence in the absence of some voluntary act that evidences the private mental states that we need to know about in order to know what kind of punishment is deserved. On this view, it is because morality requires punishers to have adequate evidence of facts that are difficult to evidence—namely mental states—that the VAR is justified. Call this the "Evidentialist Rationale" for the VAR.

The Evidentialist Rationale provides a satisfactory explanation for our allowance of criminal liability for omissions, even if they fall short of voluntary actions. What a person does *not* do sometimes tells us all we need to know about his desert-relevant mental states. It should be no surprise that the *absence* of sufficient concern for others, for instance, could be properly evidenced by the *absence* of action, at least in some circumstances; there are things that people who are sufficiently concerned would *do*, and so the failure to do them can evidence the absence of sufficient concern.

However, precisely what makes it possible to explain the law's treatment of omissions under the Evidentialist Rationale undermines the case for thinking that that rationale captures the relevant moral principle enshrined by the VAR. Omissions can provide satisfactory evidence of desert-relevant mental states *even if omissions are not voluntary acts*. In fact, there are many sources of adequate evidence of such mental states other than voluntary acts. Involuntary actions tell us a great deal about mental state; think of blushing. In fact, because it is much harder to prevent involuntary manifestations of mental state than voluntary manifestations, involuntary actions

sometimes provide *significantly better* evidence of mental state than do voluntary actions. Arguably, this is the rationale for the admissibility in court of "excited utterances" which would otherwise be inadmissible as hearsay (Federal Rules of Evidence, Article VIII, Rule 803(2). Because the excited utterance is involuntary it is more likely to be expressive of what the speaker actually believes than an unexcited utterance would be. If the VAR were supposed to ensure that the government does not violate the norm against punishment in the absence of adequately supported belief in guilt, the requirement would be at best an inept effort to conform the state's conduct to moral principle.

The Actual Invasion Rationale

Sometimes we should neither censure nor punish conduct deserving of such a response because we lack standing to do so. Intuitively, the wrongdoing in question is none of our business. You hear a stranger say something sexist to his wife at a nearby restaurant table. His conduct is wrongful and he would deserve to be reprimanded. But you ought not reprimand him, or demand that he apologize. In fact, not just the man, but the person he insulted with his comments, would have grounds for complaint were you to get involved. It's simply none of your business.

Broadly speaking, when it comes to criminal law, the state's business is with invasions of the legally protected interests of citizens; anything else is not its business. Add that only voluntary acts actually invade the legally protected interests of others, and we reach the conclusion that for the state to punish in the absence of a voluntary act would be for the state to violate the norm against punishing in the absence of standing to punish. Call this the "Actual Invasion Rationale" for the VAR.

To explain the fact that faultless involuntary behaviors are excluded from criminal punishment by the VAR, the Actual Invasion Rationale must be coupled with a particular conception of a legally protected interest. Imagine that D, who has no history of seizures, has one while driving and injures V. Has V's legally protected interest in bodily security been invaded? Under some conceptions of a legally protected interest, it has. But we might think that the legally protected interest with which the criminal law concerns itself is the interest in not being *wronged*, and not just the interest in not being *harmed*. V in this example has been harmed, but since D was in no way at fault for the harm that he caused, V has not been wronged.

Even when "invasions" are understood as "wrongings", the Actual Invasion Rationale falters with respect to two features of the law identified in section 1, namely, the exclusion from criminal liability of behaviors guided by non-conscious mental representations, and the inclusion of habitual acts. First, there is no reason to think that a victim is wronged by a harmful bodily movement only if the mental state guiding it is conscious. The officer Newton killed was wronged by the killing. If we hesitate to accept that conclusion, it is because we have doubts about the mental state that accompanied Newton's bodily movements when he pulled the trigger. We wonder, for instance, if he really intended to kill given his state of shock following his injury. But to take such intuitions to support the claim that the officer was not wronged is to fail to distinguish a lower grade of wrongdoing from its absence. Perhaps the officer was less badly wronged if Newton killed him with some mental attitude falling short of intention, or as a result of provocation, or emotional disturbance. These considerations point to differences in degrees of wrongdoing that the laws against homicide take into consideration in grading the crime. But it is to succumb to confusion to suggest that no matter what Newton's non-conscious mental state, the officer was not wronged since

Newton was not conscious. The bodily movements in which Newton engaged are *potentially* wronging; they would wrong the officer were they accompanied by objectionable mental states, and so they cannot be excluded, wholesale, from criminal liability under the Actual Invasion Rationale. They *did* actually invade the officer's legally protected interest, even if it is unclear how bad the invasion was.

On the flipside, without knowing whether habitual actions are either willed, or result from prior fault, we are lacking any grounds on which to insist that those who are harmed by them are wronged. If to be injured by a habitual action is no different, substantively, from being injured by a falling tree limb, broken by the wind, there is no sense in which such injuries wrong their victims. The result is that to allow criminal liability in such cases in the absence of any showing to the effect that there is either a willing that issued in the bodily movement or prior fault of some sort is to fail to conform to the moral principle that is taken to be crucial by the advocate of the Actual Invasion Rationale.

The PAP Rationale

One might think that to punish in the absence of a voluntary action is to punish where the agent could not have done otherwise, or was not free. Add that the ability to do otherwise is necessary to be justifiably held responsible, and it seems that the VAR has a natural rationale. To pursue this line is to take the VAR to be the law's way of concretizing an appealing moral principle, sometimes called "The Principle of Alternate Possibilities" (PAP), according to which *a person is morally responsible for his conduct only if he could have done otherwise*. Call this the "PAP Rationale" for the VAR.

PAP itself has come under attack since Harry Frankfurt's important paper offering counterexamples to it (Frankfurt (1988)). Frankfurt constructed examples of people who seem unable to do otherwise because of something that would block them from doing otherwise were they to try to, but never try to. Since the people in the examples just act badly for their own reasons, without ever trying and failing to do otherwise, they seem to be responsible for their bad behavior. If PAP is false, then that would indeed undermine its power to justify a legal practice that aims to induce legal actors to comply with it. However, even if we grant the truth of PAP, the PAP Rationale is inadequate. Note, first, that defendants who engage in voluntary actions need not be shown to have had the ability to do otherwise in order to be held guilty. This point would not in itself impugn the PAP Rationale were showing that a person acted voluntarily sufficient for showing that he had the ability to do otherwise. But it is not. As has long been recognized, voluntariness, while necessary for freedom, is not sufficient for it. John Locke offered a convincing counterexample to the sufficiency claim in his *Essay Concerning Human Understanding* in the late 17th century:

> [S]uppose a Man be carried, whilst fast asleep, into a Room, where is a Person he longs to see and speak with; and be there locked fast in, beyond his Power to get out: he awakes, and is glad to find himself in so desirable Company, which he stays willingly in, *i.e.* preferrs his stay to going away. I ask, Is not this stay voluntary? I think, no Body will doubt it: and yet being locked fast in, 'tis evident he is not at liberty not to stay, he has not freedom to be gone. (Locke (1979))

Imagine that the man in the example lacks permission to stay and so is trespassing, which is a crime. While we can argue over whether it is just to punish him for staying,

the VAR *itself* does not provide us with reason to object; the man engaged in a voluntary act; his bodily movement, in staying in the room, was willed. There would be nothing objectionable by the standards of the VAR about a statute that defined a crime in such a way that this man's conduct met the definition, as statutes prohibiting trespassing do.

A response to this concern can be made through noting that cases of voluntariness without the ability to do otherwise are rare. But this amounts to the assertion that it is rare under the VAR to punish someone who is unworthy of punishment. If the ability to do otherwise is what we really require for criminal liability, then wrongful conviction would be even rarer were we to write PAP into law.

The PAP Rationale also yields the wrong results about habitual action. Do those who engage in habitual actions have the ability to do otherwise? Under one construal, habitual actions are simply "triggered" by perception of the environment. It is because he hears something threatening behind him that the soldier, in the earlier hypothetical, spins and fires. But given that it is not in his control that he should hear the threat, it is not under his control that he should spin and fire on the person who issued it. Were the PAP Rationale correct, we would have good reason to exclude habitual actions from criminal liability, rather than include them, as we do.

The Reasons-Responsiveness Rationale

Those who accept the arguments against PAP, and so hold that it is possible to be fully morally responsible for one's conduct even in the absence of the ability to do otherwise, might nonetheless hold that there is a necessary condition of responsible conduct that is not met when the agent being assessed has not performed a voluntary

act. According to John Fischer's position, a person is morally responsible for his behavior only if it issued from a "reasons-responsive" mechanism. (See, for instance, the essays collected in Fischer (2006).) While the full set of conditions that must be satisfied for a mechanism giving rise to behavior to count as reasons-responsive, in Fischer's sense, is complex, the basic idea is this: *responsibility requires that the agent would have recognized and responded to reasons to do otherwise had there been such reasons*. It is possible to be responsive to reasons for alternatives even while lacking the ability to pursue them: imagine that the man in the locked room would have tried to leave were he offered a sum of money; so he would have responded to reasons to leave, but still would have stayed. Since reasons-responsiveness and the ability to do otherwise pull apart, Fischer's condition on responsibility is distinct from PAP.

Fischer's idea suggests an alternative rationale for the VAR: When an agent's behavior is not voluntary, by the law's standards, it arises from a process, or a mechanism, that is not reasons-responsive. Therefore, if reasons-responsiveness is necessary for responsibility, and responsibility is necessary for desert of punishment, the Voluntary Act Requirement helps the law conform to moral principle. Call this the "Reasons-Responsiveness Rationale".

This rationale succeeds in accounting for many aspects of the VAR. Martin, who was dropped on the highway by the police and then arrested for appearing there, did not appear on the highway as a result of his own reasons-responsive mechanism; no matter what reasons he had and recognized for not appearing on the highway, he was going to appear there given the police's determination that he should. A defendant who flails his arms while in the grips of a seizure, for the same reasons, does not engage in bodily movements that have their source in a reasons-responsive mechanism; the electrical activity that gives rise to the relevant bodily movements would not alter in the

face of reasons not to move the body in that way. And the Reasons-Responsiveness Rationale does well in accounting for the fact that omissions are as good as voluntary acts for legal purposes: a person who omits to aid her child, for instance, might very well fail to do so as a result of the workings of a mechanism such that, had she had certain reasons for helping her child, she would have recognized and responded to them. Maybe, for instance, she would have done so had someone offered her crack in exchange. If so, then that fact, on Fischer's view, would show that she omitted through a reasons-responsive mechanism. (Fischer and Ravizza (1988), ch. 5).

The Reasons-Responsiveness Rationale does substantially less well at explaining some other aspects of the VAR. First, consider the law's exclusion of bodily movements guided by non-conscious mental states from the category of the voluntary. There is no reason to suppose that consciousness is required for reasons-responsiveness. Newton, for instance, seemed to be acting from a reasons-responsive mechanism when he ran from the scene of the shooting to the hospital: had he thought the hospital was elsewhere, he would have run there instead. And yet he is not to be held criminally liable for his bodily movements. Either this is an error on the law's part, or else the Reasons-Responsiveness Rationale falls short of identifying the moral principle underlying the VAR.

Further, consider the fact that habitual actions are to be treated as though they were voluntary actions. Willed or not, such bodily movements do not arise from a reasons-responsive mechanism. Consider again the soldier whose habitual reflexes leading him to shoot someone at a firing range whose threatening conduct warrants a much less aggressive response, if any at all. The circumstances that "trigger" the defendant's bodily movement—the victim shouting an obscenity towards him—would trigger that bodily movement no matter what reasons the defendant had not to squeeze

the trigger. In fact, he had plenty such reasons and failed to respond to them. The mechanism in this case simply does not respond to reasons in the way that one might take to be necessary for moral responsibility.

Some will be drawn to the idea that the soldier, in the hypothetical just offered, ought not to be going to the firing range knowing, as he does, of his conditioned, reflexive, habits. That is, some will be attracted to the possibility of assigning some kind of responsibility in this case through appeal to prior fault—in this case prior voluntary action that placed the agent in a situation where he would be at risk of encountering a "trigger" of his habit of shooting at threats behind him in circumstances in which he would be armed. However, the law does not inquire in habitual action cases about prior fault. The defendant in a case like this would need not be shown to have been at fault for being in the circumstances in which he found himself, nor would the behaviors that got him there need to be shown to have been voluntary. The mere fact, rather, that the relevant bodily movements were the product of a habit is sufficient to show there to be compliance with the VAR in assigning a guilty verdict. For instance: imagine that the defendant believed, and had every reason to believe, when he went to the firing range that he would handle only unloaded weapons. In that case, he could not be guilty of the crime through appeal to his earlier voluntary acts leading to his presence at the firing range; at the time of those bodily movements, he had no reason to believe that he was even risking harm to anyone else of the sort that he eventually caused. However, under the criminal law, the case would need to be made that he did not know he was risking harm to anyone *at the time of the habitual bodily movements*; his earlier mental state would be relevant only in so far as it provided evidence of his later mental state. In stipulating that habitual bodily movements are to be treated as voluntary

actions, the criminal law is not pursuing the kind of prior fault analysis to which an advocate of the Reasons-Responsiveness Rationale might be attracted.

3. Enacting the Requirement of Correspondence

When a defendant is shown to be guilty of a crime, he is shown to have performed certain acts with certain results in certain circumstances—he is shown, for instance, to have lit a fire (an act) that burned a piece of property (a result) that was not his own (a circumstance). And he is shown to have been in certain mental states—he is shown, for instance, to have intended to light the fire, to have been aware of a substantial and unjustifiable risk that property would burn as a result, and to have known that the property in question was not his own. But, in fact, there is an additional requirement that is so rarely at issue as to go unmentioned most of the time: the defendant's actions must correspond with his mental states. The fact that the defendant once long ago intended to light the fire matters not at all if he didn't intend to light the fire that he lit *when he lit it* or if he didn't light it *as a result of* that very intention. The mental states that matter must be linked to ("concur with", "correspond to") the acts that form the basis of criminal liability. Call this the "Requirement of Correspondence".

Under what conditions do the act and mental state "correspond"? Criminal law students are taught that the needed correspondence is present when the act and the mental states take place at the same time. Teaching the Requirement of Correspondence this way may serve pedagogical purposes, but it misses the point of the requirement. We can illustrate this with examples in which an agent has two intentions at a particular time, but is only acting on one of them. D and V go hunting. D intends to kill V when the moment is right, but intends to hunt deer until then. When

he fires towards both a deer and V, and hits V, killing him, is there appropriate correspondence between his intention to kill V and his act? There is not if at that moment he was acting on his intention to kill the deer and merely accidentally killed V. Of course, such a defendant has an uphill battle to convince a jury that he was not acting on his intention to kill V. But in fighting it he is trying to show that the Requirement of Correspondence was not met. The Requirement of Correspondence does not merely bar liability when the act and mental state are not simultaneous. Rather, it bars liability when the act is not *the manifestation* or *the product* of the mental state. In fact, there is an intuitively appealing moral principle here: *A person's mental states contribute to his responsibility for his act only if the act is the manifestation (in an intuitive but not yet defined sense) of those mental states*. If this is right, then a person is morally responsible for a package consisting of mental states and conduct only if the mental states in question are manifested in the conduct. The parts of the package, that is, must be glued together, as it were, for there to be responsibility for the package.

It is a foundational principle of criminal law that part of what justifies punishing a defendant are the objectionable mental states that he was shown to have had. Further, when we grade a crime on the basis of mental state—giving, for instance, a lower penalty for a negligent than for an intentional homicide—we assume that the moral quality of the act is a function in part of the mental state that gave rise to it. However, if the moral principle about correspondence is correct, then mental states contribute to the defendant's responsibility only if they are manifested in action. The homicide, for instance, deserves to be classified as intentional, rather than negligent, that is, only if the intention to kill was manifested in the act of killing. But, for reasons to be explained in this section, mental states are manifested in action (at least in the paradigm case) only if the defendant performed a voluntary act. Hence, the VAR is justified because failure to

comply with it would involve failure to comply with the moral principle concerned with correspondence. Voluntary acts matter to criminal liability, on this view, because without them we lack the link between objectionable mental states and objectionable acts that is required to be justified in punishing for the package of mental states and conduct that crimes, in fact, consist in. There is *mens rea* and there is *actus reus*; but without a voluntary act, there is not the link between the two that is required for desert of punishment for the conjunction.

Call this the "Manifestation of Mens Rea Rationale" for the VAR. Under it, the VAR is a byproduct of the idea that *mens rea* is an essential part of criminal liability. It is because we already think that people should not be punished in the absence of a showing of *mens rea*, under the rationale, that we are barred, for moral reasons, from punishing them in the absence of a voluntary act. *Mens rea* is essential, but it isn't relevant unless it's manifested. And it isn't manifested unless there's a voluntary act. To punish, then, in the absence of a voluntary act is morally no different from punishing in the absence of *mens rea*; and that is unacceptable. The rest of this section will be spent defending this rationale for the VAR.

To see why a voluntary act is required for there to be correspondence with mental state and conduct, consider, first, the paradigm case: At t1, D decides to kill V and forms an intention to do so; at t2, acting on that intention, D squeezes the trigger of a gun and V is killed. In this description of the case, the crucial phrase is "acting on that intention". To say that is to say that the Requirement of Correspondence has been met. But what does that phrase mean? Under what conditions is a person acting on a particular intention? Consider the following possible answer:

At t2, D acts on his intention to kill V *if and only if* The intention causes D's bodily movement (the movement of his finger on the trigger) at t2.

As attractive as this simple theory is, it will not do. The problem is well-known to philosophers of action: if the causal route from intention to bodily movement is "deviant" then the right side of the biconditional will be satisfied, but not the left. For instance, say that at t2 D is at the firing range and notices that V is passing between him and the target at which he is pointing. Recalling that he earlier formed an intention to kill V, D becomes very nervous—"My God," he thinks, "I might actually go through with this!"—and his nervousness causes his hands to shake violently, resulting in the bodily movement of his finger pulling the trigger. In such a case, the bodily movement is not the manifestation of the intention of the sort that is required for correspondence, or, as we say, for it to be the case that D is "acting on the intention". In that case, although the intention causes the bodily movement—the causal sequence is intention-nervousness-finger movement—the bodily movement is not the manifestation of the intention in the sense that matters for responsibility.

We can't solve the problem by insisting that there be no causal intermediaries between intention and later bodily movement when someone acts on an intention. After all, there are causal intermediaries between the intention and the relevant bodily movement in paradigm cases of acting on an intention. You intend to go to the baseball game next weekend and so buy tickets. You have no plans about how you will get from your house to the stadium. When the time comes, you walk. In walking you are acting on the intention to go to the game. But that intention did not represent those bodily movements; you had not decided how you would get to the stadium when you formed the intention. Still, *something* must have represented those bodily movements. In fact,

we know what that something is: an intention to get to the stadium by walking. *That* later intention is a causal intermediary between your intention to go to the game and your bodily movements.

This reflection on the paradigm case, however, makes evident an important difference between the case in which the intention misfires, as in the case in which it makes D so nervous that he squeezes the trigger of the gun, and the case in which it does not, such as the case of going to the game by walking there. There is a difference between the causal intermediaries between intention and bodily movement in the two cases. In particular, in the game case the causal intermediary is itself a representation of the act it causes: you intend *to walk* and this results in your walking. The intermediary has representational content that turns out to match the world when it causes walking. By contrast, in the misfire case the causal intermediary is just nervousness, which may represent no act at all, much less the act of squeezing the trigger. That state of nervousness could have caused D's eye to twitch rather than his finger without there having been any apparent mismatch between the mental state and the world.

Notice that, in the game case, we can also look for causal intermediaries between the intention *to walk* and the bodily movement it causes. And, in fact, we find some: when you decide to walk, you haven't decided whether to start with your left foot or your right. When you begin with your right, that bodily motion is a paradigm instance of acting on the intention to walk. *Something* must have represented that bodily movement. But it wasn't the intention to walk. So, there must have been some intermediary between the intention to walk and the movement of your right foot which represented that movement. What intermediary? Probably an intention to move your *right* foot. And we can extend the point further. When you formed the intention to start with your right foot, you didn't decide that you would step with that foot *over* the

puddle in the yard rather than into it. But *something* must have represented that longer-than-normal bodily movement. What? Answer: an intention to step a longer-than-normal distance with your right foot. In the paradigm case of acting on an intention, that is, it seems that, in principle, between *any* intention and the bodily movement it causes there is some other representational state, perhaps some other intention, that represents features of that bodily movement that are not represented by the previous intention in the sequence. There is a causal sequence of intentions, that is, that are increasingly specific, or fine-grained, in their content. The first, in this case, is the intention to go to the game; later there is the intention to walk; and later still there is the intention to take an extra-long step with the right foot. It is thanks to the fact that there is this sequence ending with the movement of the foot itself that that movement is an instance of acting on the first intention in the sequence, namely the intention to go to the game.

Of course, at some point this search for causal intermediaries that represent features of the bodily movement not represented by the most recent intention we've identified will give out. Your foot lands exactly .4 inches past the edge of the puddle. But no mental state represented *that* feature of the bodily movement; it is not as though you would have failed to do as intended had your foot landed .5 inches past the puddle's edge instead. So, at some point, the mental state that serves as intermediary in the paradigm case will be as fine-grained in its content as it gets. It is an empirical question how fine-grained such representations are. Still, at least in the paradigm case, there must be causal intermediaries with remarkably fine-grained content: in stepping over the puddle—in stepping three feet rather than merely two—you are acting on your intention *to go to the game*, and so there must be some mental state, caused by the intention to go to the game (among other things, such as a belief that the puddle is in

the way), that represents a longer-than-normal bodily movement and thus causes it to be the case that you engage in one. Let's give the name "volition" to the mental state that is as fine-grained in content as it gets and which we find in the paradigm case of acting on an intention. (John Searle uses the term "intention-in-action" to refer to this mental state. See Searle (1983), ch.3.)

The central point here is, perhaps, already clear: if all cases were paradigm cases, then whenever there is appropriate correspondence between intention and later bodily movement—correspondence of the sort that is needed for the bodily movement to be the manifestation of the intention in the sense that matters for responsibility for the intention-action package—the bodily movement would have to be caused by a representation of it, which is part of what is needed under the VAR. In paradigm cases of legal importance, the relevant intention is not the trivial intention to move the body in a certain way. It is, instead, an intention that is not nearly so fine-grained in its content, such as an intention to kill, or an intention to remain somewhere that you have no right to remain, or the intention to deface public property. What we need to know is whether the bodily movements that then cause a death (in homicide), or cause someone to remain where they shouldn't (in criminal trespass), or cause public property to be defaced (in destruction of public property) were manifestations of these objectionable intentions. In the paradigm case, these bodily movements were not such manifestations if they weren't guided, at the time, by mental representations of them, volitions, in the sense just defined. And so, since we care about the Requirement of Correspondence, we need to care about guidance of bodily movement by volition.

What has been said so far, then, explains the defining feature of the VAR, namely that, putting aside exceptions, there is no criminal liability in the absence of a bodily movement that is immediately guided by a mental representation of it. But can appeal

to the Manifestation of Mens Rea Rationale explain the features of the VAR identified in section 1? Yes.

First, consider the law's refusal to allow bodily movements guided by non-conscious representations to count as voluntary acts for the purposes of the VAR. To see this that this can be explained, first consider the Model Penal Code's definition of an intention (or "purpose") to act or bring about a result:

> A person acts purposely…if…it is his conscious object to engage in conduct of that nature or to cause such a result. (MPC §2.02(2)(a))

By requiring that a legally-relevant intention be *conscious*, as specified here, the Model Penal Code sets the rule that we never impose criminal liability for unconscious intentions. The framers of the Code may be concerned to exclude conviction for purposeful crimes through appeal to Freudian explanations. The accidental killing of the defendant's father, for instance, can be made to appear to be a purposeful killing through assignment to the defendant of a subconscious intention to kill his father; add an Oedipal diagnosis and a jury could be led to a very draconian verdict. If we think the requirement that the intention be conscious is a good one, the question then is whether, in the paradigm case, the causal intermediaries between conscious intention and bodily movement can be *un*conscious when the bodily movement is a manifestation of the intention. If the volition were unconscious, that would at least provide reasonable doubt as to whether the act were really the manifestation of the conscious intention in the way that matters for criminal responsibility. The sufferer from the Oedipal complex who has managed to bring his intention to kill his father into consciousness, might nonetheless repress into subconsciousness the mental states

through which he executes that conscious intention. To convict in such a case for intentional homicide on proof of the intention to kill would be no different, in the end, from convicting for intentional homicide when the intention to kill itself is subconscious. Therefore: it is because we require conscious intention that we require conscious volition under the VAR. Without consciousness of volition, conscious intention would not be properly manifested in action. In a case like Newton, then, there are really two reasons not to ascribe Newton with criminal responsibility for the killing: his intention to kill (if he had one) was unconscious, and so were the mental states through which that intention became manifested in the bodily movements that caused the death.

Second, consider the rule that a certain class of omissions can suffice for the VAR. This practice too admits of an explanation under the Manifestation of Mens Rea Rationale: what we *do not do* often manifests our objectionable mental states just as much as what we *do do* even if there is no volition present. The mental state of disregarding one's child's welfare is manifested by *the failure* to do that which the child's welfare requires that one do. However, in such cases, there need be no volition to serve as causal intermediary between the morally and criminally relevant mental state—in the imagined case the state of "disregard"—and the failure to do as one ought. While that failure may need to be caused by the prior objectionable mental state for the failure to manifest it in the morally relevant way, such causation does not require volitional intermediaries.

And, third, consider the case of habitual actions. The rationale for treating convictions for habitual actions as complying with the VAR is, roughly, that we have reason to believe that, like some omissions, some habitual actions can be manifestations of objectionable mental states in the sense that matters to morality, even if they are not

guided by conscious volitions. One benefit of a habit is that it produces conduct for which there are reasons without the agent taking the time to reflect on and weigh those reasons; the habit of rolling through a stop sign on the way to work allows D to shave a few seconds off his commute while, at the same time, not interrupting his musings about how to handle his boss that day. However, a byproduct of this valuable feature of habits is that they override our tendencies to withold action in the face of reasons to do so. Since D has a habit of rolling through the stop sign, he does so on a particular occasion even though consciously aware of a much greater danger on that occasion than usual. In this case, the mental state of recklessness manifests itself in the bodily movement *thanks to* the fact that the bodily movement is habitual. Normal agents find themselves in objectionable mental states all the time; but most of us override the impulses to conduct that those mental states supply. Failure to override is one way in which such mental states can be manifested in behavior, and this is found in at least some habitual actions. Hence to exclude them on the grounds of involuntariness would be to miss the point of the VAR. It is because that requirement helps us to be sure to comply with the Requirement of Correspondence that habitual actions, also, serve for criminal liability.

One objection to the Manifestation of Mens Rea Rationale arises from reflection on cases of "strict liability": cases in which guilt can be established despite the absence of a showing of any particular culpable mental state with respect to crucial elements of the crime's *actus reus*. Consider the case of State v. Kremer (262 Minn. 190, 114 N.W.2d 88 (1962)). Kremer's brakes failed, without any prior warning, and he ran through a flashing red light, in violation of a city ordinance. To be guilty of the crime, a defendant need not be shown to have been even negligent with respect to the fact that he was running the light. There is guilt even if, for instance, the light were covered from view

by a tree branch, so that even a reasonable person exercising due care would not have known that he was running the light. Kremer appealed his conviction on the grounds that it was in violation of the VAR. The question before the court, then, is whether the VAR applies even when the offense is strict liability.

The Manifestation of Mens Rea Rationale might appear to imply that when the crime is strict liability, as in Kremer, the VAR is irrelevant: if no mental state is required for guilt, then it seems that a requirement meant to assure that culpable mental states *be manifested* does not apply. However, the court in that case goes the other way, quashing Kremer's conviction on the grounds that it was in violation of the VAR. Indeed, this is settled law. However, it is only through a mistaken conception of strict liability that one can reach the objection to the Manifestation of Mens Rea Rationale just offered. There are two ways to construe strict liability crimes: (1) The crime has no *mens rea* requirements, or (2) The crime has *mens rea* requirements but any mental state on the part of the defendant meets them. To conceive of strict liability in the first way is to see strict liability crimes as involving a major departure from fundamental axioms of criminal law, particularly the principle according to which acts are never worthy of punishment in the absence of accompaniment by culpable mental states. While there is good reason to think this is the wrong way to construe strict liability, if it were the right way to construe it, then it would follow that the court in Kremer made a mistake: the VAR would not apply. It is hard to see how someone willing to accept strict liability on these grounds could object to this conclusion. If we are ready to overturn fundamental principles of the morality of criminal law for the sake of social order, why not the VAR too?

Under the second construal of strict liability, however, the Kremer court made the right decision, but for reasons that are compatible with the Manifestation of Mens

Rea Rationale. Under that construal, running a flashing red light is punishable only if it manifests a culpable mental state; but, as it happens, *any* mental state would be culpable. Hence, the behavior is punishable only if it manifests *some* mental state. The trouble with convicting Kremer, however, is that his act manifested *no mental state at all*. Under the second construal of strict liability, that is, the Requirement of Correspondence still needs to be met. And so the VAR still needs to be met, under the Manifestation of Mens Rea Rationale for it.

Conclusion

The Manifestation of Mens Rea Rationale explains central features of the legal doctrines involved in the VAR that are either inexplicable, or explicable inelegantly, through appeal to alternative rationales. What this suggests is that the central point of the VAR is to enshrine in law a common sense moral requirement: we are not morally responsible for a package of objectionable mental states and harmful conduct unless the conduct is the manifestation of the mental states. This is a condition that is not met *in the paradigm case* in the absence of a voluntary act. And when the case differs from the paradigm case, and the condition is nonetheless met, as in some cases of omission and habitual action, the VAR allows criminal liability in the absence of a voluntary act.

So, in so far as criminal punishment should be applied only to the morally deserving, we ought to conform, as we do, to the VAR. The VAR enshrines in law a maxim of morality as fundamental as the maxim against punishing someone other than the perpetrator of the crime. The maxim that underlies it is as fundamental—in fact, as has been argued here, *exactly* as fundamental—as the maxim against punishing in the absence of a showing of mental state. It is because the criminal's mind matters to his

responsibility that we punish him only if it has been manifested in his conduct.

Works Cited

Bach, Kent. (2010) "Refraining, Omitting, and Negative Acts" in A Companion to the Philosophy of Action, T. O'Connor and C. Sandis eds., Oxford: Wiley-Blackwell.

Fischer, John. (2006) My Way: Essays on Moral Responsibility, Oxford: Oxford University Press.

Fischer, John and Ravizza, Mark. (1998) Responsibility and Control: A Theory of Moral Responsibility, Cambridge: Cambridge University Press.

Frankfurt, Harry. (1988), "Alternate Possibilities and Moral Responsibility" in The Importance of What We Care About, Cambridge: Cambridge University Press.

Husak, Doug. (2007) "Rethinking the Act Requirement," 28 *Cardozo Law Review*, pp. 2437-2460.

Locke, John. An Essay Concerning Human Understanding, P.H. Nidditch, ed., Oxford: Clarendon Press, 1979, II.xxi.10, p. 172.

Searle, John. (1983) Intentionality, Cambridge: Cambridge University Press.

Further Reading

Michael Bratman. (1994) "Moore on Intention and Volition" in *University of Pennsylvania Law Review*, Vol. 142, No. 5, pp. 1705-1718.

Chiao, Vincent. (2009) "Action and Agency in the Criminal Law" in Legal Theory, v. 15, pp. 1-23.

Duff, R.A. (2007) Answering for Crime: Responsibility and Liability in the Criminal Law, Oxford: Hart Publishing.

Husak, Doug. (2010) "Does Criminal Liability Require an Act?" in The Philosophy of Criminal Law: Selected Essays, Oxford: Oxford University Press.

Moore, Michael. (1993) Act and Crime, Oxford: Oxford University Press.

Moore, Michael. (1997) Placing Blame, Oxford: Oxford University Press.