Strawson’s thesis says that the reactive attitudes involving praise and blame are apt if their targets are responsible. The realist interpretation of Strawson’s thesis claims that it is the responsibility of the targets that grounds and explains the fittingsness of the reactive attitudes toward these targets. The retributivist is also a realist about the relationship between responsibility and the reactive attitudes and practices of blame and punishment -- culpable or responsible wrongdoing is the desert basis of blame and punishment. We need to see if there is a plausible conception of responsibility to ground blame and punishment.

KINDS OF RESPONSIBILITY
We talk about responsibility in different contexts with different presuppositions and commitments. We can distinguish at least three distinct kinds of responsibility.

Sometimes when we say that A was responsible for x we signal that A was causally responsible for x. In such cases, A is a sufficient cause of x. But we can see that in cases of causal responsibility the cause need not be an agent of any kind. A hailstorm might be responsible for the crack in your windshield. Causal responsibility does not implicate blame or punishment.

We should focus on the way in which agents might be responsible for outcomes. Writing primarily about moral responsibility, Gary Watson has distinguished two faces of responsibility: attributability and accountability. As a rough first approximation, we ascribe responsibility in the attributive sense to agents for their actions insofar as we see those actions reflect the quality of their wills, whereas we ascribe responsibility in the accountability sense to agents for their actions insofar as it is fair to blame them for those actions.

We hold an agent attributively responsible for her actions and the foreseeable consequences of those actions based on the quality of her will, where that discloses her character or true self and reflects her fundamental evaluative orientation. We might not hold an agent responsible for harm that she causes to others if this was beyond her control, for instance, if she was manipulated by natural forces or the will of another. But we do hold someone responsible in one way if the harms she causes reflect her will in certain ways, for instance, if she intended the harm or was aware of the risks she posed to others and was indifferent. Different conceptions of quality of will are possible. We might understand it (1) in characterological terms as a will expressing one’s stable character traits, (2) in terms of a will in which the agent endorses her motivating desires, or (3) in terms of the regard and concern the agent displays for the interests and rights of others. However, it is not clear that we should limit responsibility and blame to culpable states of mind that express an agent’s stable character traits or dispositions that she

---


2 Hume gives expression to a characterological conception of responsibility and quality of will in David Hume, An Enquiry Concerning the Principles of Morals (1751), §VII, Part II.

endorses, for then we could not hold agents responsible for wrongs that were out of character or for familiar forms of weakness of will, in which the agent acts on desires she does not endorse. It seems more reasonable to identify an agent’s quality of will with the kind of regard she has for the interests and rights of others.

T.M. Scanlon develops a conception of blame that appeals to attributive responsibility, where he understands quality of will in terms of the regard that the agent shows for the interests and rights of others. Like Strawson, Scanlon thinks that reactive attitudes, such as blame, reflect our relationships with others and that blame is the appropriate response to an unjustified breach in these relationships. In assessing such breaches, Scanlon thinks we attach significance to the motives and intentions of the agent. On this view, A is justified in her hard feelings toward B and in blaming B if B injures A through malice, indifference, recklessness, or negligence. Here, our reactive attitudes are supposed to track the insufficient regard that B shows A’s interests and rights.

There may be senses of responsibility and blame that track attributability. But notice that agents might be responsible in this attributive sense without being morally responsible. Moral responsibility and blame -- accountability -- requires more. Consider again a case in which B intends to violate A’s rights to bodily integrity or property. We said that this licenses us in attributing harmful agency to B and would justify A’s anger and hard feelings toward B. Perhaps this would even license a kind of blame toward B. But suppose that we learn that B did not in any relevant sense have a fair opportunity to do otherwise, either because B was insane and lacked basic normative competence to recognize or conform to the relevant norms or because B’s behavior was the product of duress in which a third party threatened B or his loved ones with grievous bodily harm if he did not violate A’s rights. Insanity and duress are excuses and imply that the agent was not at fault for the harm he did. It is natural in such circumstances to say that B was not responsible for the wrongs he committed and to withhold blame. But since by hypothesis B was attributively responsible for the harm he caused, the sense in which he is not responsible must be a different sense of responsibility. This is Watson’s sense of responsibility as accountability. It is also natural to say in these circumstances that B is not blameworthy for his wrongdoing. It would certainly be unfair to punish B for wrongdoing that he did not have an adequate opportunity to avoid. But blame is itself a sanctioning response, even if it is usually less severe than punishment, and so it would be unfair to blame B for wrongdoing that he did not have adequate opportunity to avoid. This means that retributive blame and punishment presuppose accountability and not just attributability. Because the fairness of blaming and punishing wrongdoing presupposes responsibility as accountability, we will focus on responsibility as accountability. Indeed, unless otherwise noted, this is how we will understand responsibility throughout the rest of the course.

RESPONSIBILITY AND EXCUSE

Predominant retributivism claims that culpable or responsible wrongdoing is necessary for blame and punishment and sufficient to establish a pro tanto case for blame and punishment. As we saw, predominant retributivism has considerable intuitive appeal and fits well with important parts of criminal jurisprudence. If blame and punishment are predicated on wrongdoing for which the agent is culpable or responsible, as the predominant retributivist claims, this explains well the two main kinds of affirmative defense an agent might offer when blame or punishment threatens --

---


5 Of course, while blaming and punishing normatively incompetent wrongdoers might be unfair, civil commitment might nonetheless be appropriate.

6 The possibility of wrongdoing for which the agent is not responsible because she lacked the fair opportunity to avoid wrongdoing is perhaps the best reason for rethinking the voluntarist claim that ought implies can. But that is the topic for another paper.
Justifications and excuses. Justifications deny wrongdoing, whereas excuses deny culpability or responsibility. Just as justification is the flipside of wrongdoing, so too excuse is the flipside of responsibility. These defenses are part of the criminal law but also reflect well the moral landscape. When others threaten to blame us morally, it can be appropriate to respond by citing factors that would justify or excuse our behavior.

Because the denial of culpability in this sense is an excuse, culpability and excuse are inversely related. Culpability and excuse should have corresponding structure, and either could be studied by studying the other.

TWO KINDS OF EXCUSE

When we adopt this methodology, we might factor responsibility into two main conditions: normative competence and situational control. This conception of responsibility draws on ideas in the reasons-responsive wing of the compatibilist tradition of thinking about free will and responsibility and ideas from the fair choice approach to criminal responsibility.\(^7\)

If someone is to be culpable or responsible for her wrongdoing, then she must be a responsible agent. Our paradigms of responsible agents are normal mature adults who are normatively competent. They must be able to distinguish between the intensity and authority of their desires and impulses. This requires that agents not simply act on their strongest desires, but be capable of stepping back from their desires, evaluating them, and acting for good reasons. If so, normative competence involves reasons-responsiveness, which itself involves both cognitive capacities to distinguish right from wrong and volitional capacities to conform one's conduct to that normative knowledge.

It is important to frame this approach to responsibility in terms of normative competence and the possession of these capacities for reasons-responsiveness. In particular, responsibility must be predicated on the possession, rather than the use, of such capacities. We do excuse for lack of competence. We do not excuse for failures to exercise these capacities properly. Provided the agent had the relevant cognitive and volitional capacities, we do not excuse the weak-willed or the willful wrongdoer for failing to recognize or respond appropriately to reasons. If responsibility

\(^7\) While justifications and excuses are the two main affirmative defenses available to defendants, there are also some policy-based exemptions, such as prosecutorial immunity for diplomats. See, e.g., Paul Robinson, Structure and Function in Criminal Law (Oxford: Clarendon Press, 1997), pp. 96-124, 204-07, and Mitchell Berman, “Justification and Excuse, Law and Morality” Duke Law Journal 53 (2003): 1-77. We might divide defenses into four kinds: (1) no violation, (2) justified or permissible violation, (3) nonculpable or blameless violation, and (4) non-punishable violation. Because (1) denies that the offense has occurred, it is not a true affirmative defense. (4) is an affirmative defense, but represents pragmatic, rather than principled, grounds for the defense. Robinson makes a good case that the communicative functions of a criminal code would be better served if acquittals reflected these alternative rationales more clearly.

were predicated on the proper use of these capacities, we could not hold weak-willed and willful wrongdoers responsible for their wrongdoing. Indeed, the fact of wrongdoing would itself be exculpatory, with the absurd result that we could never hold anyone responsible for wrongdoing. It is a condition of our holding wrongdoers responsible that they possessed the relevant capacities.

Normative competence, on this conception, involves two forms of reasons-responsiveness: an ability to recognize wrongdoing and an ability to conform one's will to this normative understanding. Both dimensions of normative competence involve norm-responsiveness. As a first approximation, we can distinguish moral and criminal responsibility at least in part based on the kinds of norms to which agents must be responsive. Moral responsibility requires capacities to recognize and conform to moral norms, including norms of moral wrongdoing, whereas criminal responsibility requires capacities to recognize and conform to norms of the criminal law, including norms of criminal wrongdoing.

Normative competence requires the cognitive capacity to make suitable normative discriminations, in particular, to recognize wrongdoing. If so, then we can readily understand one aspect of the criminal law insanity defense. Most plausible versions of the insanity defense include a cognitive dimension, first articulated in the M'Naghten rule that excuses if the agent lacked the capacity to discriminate right from wrong at the time of action.9

But there is more to normative competence than this cognitive capacity. It is a common view that intentional action is the product of informational states, such as beliefs, and motivational states, such as desires and intentions. Though our beliefs about what is best can influence our desires, producing optimizing desires, our desires are not always optimizing. Sometimes they are good-dependent but not optimizing, when they are directed at lesser goods, and sometimes they are completely good-independent. This is reflected in cases of weakness of will in which we have beliefs about what is best (and perhaps optimizing desires) but in which we act instead on the basis of independent non-optimizing passions and desires. This psychological picture suggests that being a responsible agent is not merely having the capacity to tell right from wrong but also requires the capacity to regulate one's actions in accordance with this normative knowledge. This kind of volitional capacity requires emotional and appetitive capacities to enable one to form intentions based on one's practical judgments about what one ought to do and execute these intentions over time, despite distraction, temptation, and other forms of interference.

If one's emotions and appetites are sufficiently disordered and outside one's control, this might compromise volitional capacities necessary for normative competence. Consider the following obstacles to volitional competence.

• Irresistible desires or paralyzing fears that are neither conquerable nor circumventable, as perhaps in some cases of genuine agoraphobia or addiction.10
• Clinical depression that produces systematic weakness of will in the form of listlessness or apathy.
• Acquired or late onset damage to the prefrontal cortex of the brain in which agents have considerable difficulty conforming to their own judgments about what they ought to do, as in the famous case of Phineas Gage.11

9 M'Naghten's Case, 10 Cl. & F. 200, 8 Eng. Rep. 718 (1843).
10 Cf. Al Mele, “Irresistible Desires” Nous 24 (1990): 455-72. A desire is conquerable when one can resist it and circumventable when one can perform an action that makes acting on the desire impossible or at least more difficult. The alcoholic who simply resists cravings conquers his impulses, whereas the alcoholic who throws out his liquor and stops associating with former drinking partners or won't meet them at places where alcohol is served circumvents his impulses. Conquerability is mostly a matter of will power, whereas circumventability is mostly a matter of foresight and strategy.
Each of these cases involves significant volitional impairment in which agents experience considerable difficulty implementing or conforming to the normative judgments they form. Recognition of a volitional dimension of normative competence argues against purely cognitive conceptions of insanity, such as the M’Naghten test, which recognizes only cognitive deficits as the basis for insanity, and in favor of a more inclusive conception, such as the Model Penal Code’s conception of insanity as involving significant impairment of either cognitive or volitional competence (§4.01).¹²

Both cognitive and volitional competence involve sensitivity to reasons. But sensitivity is a scalar notion. This raises the question how responsive someone needs to be to be responsible. This is an important and difficult issue. We might begin by distinguishing different grades of responsiveness.¹³ We can specify the degree to which an agent is responsive to reasons in terms of counterfactuals of what she would believe or how she would react in situations in which there was sufficient reason for her to do otherwise.¹⁴ An agent is more or less responsive to reason depending on how well her judgments about what she ought to do and her choices would track her reasons for action. Initially, we might distinguish two extreme degrees of responsiveness.

- **Strong Responsiveness:** Whenever there is sufficient reason for the agent to act, she recognizes the reason and conforms her behavior to it.
- **Weak Responsiveness:** There is at least one situation in which there is a sufficient reason to act, and the agent recognizes that reason and conforms her behavior to it.

However, it does not seem plausible to model normative competence in terms of either strong or weak responsiveness.

Strong responsiveness is too strong. We do not require that people actually act for sufficient reasons; it is the capacities with which they act that matter. The weak-willed are, at least typically, responsible for their poor choices. Indeed, since strongly reasons-responsive agents always recognize and conform to reasons, they would never act wrongly. But that would make wrongdoing per se excusing, which would prevent us from ever holding wrongdoers responsible.

Moreover, weak responsiveness seems too weak. It treats someone as responsive in the actual situation even if she did not respond in the actual situation and there is only one extreme circumstance in which she would recognize and respond to reasons for action. Suppose that the

---

¹¹ Phineas Gage was a nineteenth century railway worker who was laying tracks in Vermont and accidentally used his tampering iron to tamp down a live explosive charge, which detonated and shot the iron bar up and through his skull. Though he did not lose consciousness, over time his character was altered. Whereas he had been described as someone possessing an “iron will” before the accident, afterward he had considerable difficulty conforming his behavior to his own judgments about what he ought to do. The story of Phineas Gage is discussed in Antonio Damasio, *Descartes’ Error: Emotion, Reasons, and the Human Brain* (New York: Putnam, 1994).

¹² American Law Institute, *Model Penal Code*, §2.09. The Model Penal Code is a model statutory text of fundamental provisions of the criminal law, first developed by the American Law Institute in 1962 and subsequently updated in 1981. The MPC was intended to serve as a model for local jurisdictions drafting and revising their criminal codes.

¹³ Here, I adapt some ideas from Fischer and Ravizza, *Responsibility and Control*, chs. 2-3. Whereas they defend an asymmetrical conception of reasons-responsiveness that requires moderate cognitive competence and only weak volitional competence, I treat these two aspects of normative competence symmetrically and require moderate responsiveness in each.

¹⁴ In specifying an agent’s capacities in terms of such counterfactuals, I remain agnostic about whether capacities or counterfactuals have explanatory priority, in particular, whether capacities ground the counterfactuals or whether the capacities just consist in the truth of such counterfactuals.
Agoraphobe’s phobia is so irresistible that she could only resist it under imminent threat of death. Should she be regarded as normatively competent in situations involving lesser threats? For instance, should she be held responsible for not leaving her house if this was the only way to prevent a minor property crime? Weak responsiveness seems too weak.

The Goldilocks standard of responsiveness evidently lies somewhere between these extremes. Of course, there is considerable space between the extremes – the gap between always and once. We might stake out an intermediate space of responsiveness.

- **Moderate Responsiveness**: Where there is sufficient reason for the agent to act, she regularly recognizes the reason and conforms her behavior to it.

Moderate responsiveness is deliberately vague; it specifies a range or space of counterfactuals that must be true for the agent to be suitably responsive. Ideally, we would be able to specify a preferred form of moderate responsiveness more precisely. But what is important, for present purposes, is that reasons-responsiveness is a matter of degree and that the right threshold for responsibility involves some kind of regularity in tracking and conforming to one’s reasons.

An important part of an agent’s being responsible for wrongdoing that she chose and intended consists in her being a responsible agent. This we have conceptualized in terms of normative competence and analyzed into cognitive and volitional capacities. Evidence for this view is that one seems to have an excuse, whether complete or partial, if one’s normative competence is compromised in significant ways. Some of the most familiar kinds of potential excuse – insanity, immaturity, and uncontrollable urges – involve compromised normative competence.

But there is more to an agent being culpable or responsible for her wrongdoing than her being responsible and having intentionally engaged in wrongdoing. Excuse is not exhausted by denials of normative competence. Among the factors that may interfere with our reactive attitudes, including blame and punishment, are external or situational factors. In particular, coercion and duress may lead the agent into wrongdoing in a way that nonetheless provides an excuse, whether full or partial. The paradigm situational excuse is coercion by another agent, as when one is threatened with physical harm to oneself or a loved one if one doesn’t participate in some kind of wrongdoing, for instance, driving the getaway car in a robbery. The Model Penal Code adopts a reasonable person version of the conditions under which a threat excuses, namely, when a person of reasonable firmness would have been unable to resist, provided the actor was not himself responsible for being subject to duress (§2.09). Situational duress does not compromise the wrongdoer’s status as a responsible agent and does not challenge her normative competence, but it does challenge whether she is responsible for her wrongdoing.15

**RESPONSIBILITY AND THE FAIR OPPORTUNITY TO AVOID WRONGDOING**

So far, the conception of responsibility emerging here is a two-factor model twice over. Responsibility is factored into normative competence and situational control, and normative

15 The details of duress are tricky. Some situational pressures, such as the need to choose the lesser of two evils, may actually *justify* the agent’s conduct, as is recognized in *necessity* defenses. If the balance of evils is such that the evil threatened to the agent is worse than the evil involved in her wrongdoing, then compliance with the threat is justified. But in an important range of cases, coercion and duress seem not to justify conduct (remove the wrongdoing) but rather to *excuse* wrongdoing, in whole or in part. In such cases, where the evil threatened is substantial but less than that contained in the wrongdoing, the agent’s wrongdoing should be excused because the threat or pressure was more than a person should be expected to resist. Whether the difference between when duress justifies and when it excuses should be settled by applying the balance of evils test depends on the moral framework in which we measure evils. The criminal law tends to assume a consequentialist version of the test, but one might instead measure the moral seriousness of evils in ways that reflect agent-centered prerogatives.
competence is factored into cognitive and volitional capacities. This kind of two-factor model seems plausible, in significant part because it promises to fit our practices of excuse in both moral assessment and the criminal law pretty well. Perhaps this is adequate justification. But it would be nice if there were some unifying element to its structure.

One possible umbrella concept is control. Freedom from coercion and duress, cognitive competence, and volitional competence all seem to be aspects of an agent’s ability to control her actions. But control seems important, at least in part, because it seems unfair to blame agents for outcomes that are outside their control. This suggests that the umbrella concept should be fairness, in particular, the fair opportunity to avoid wrongdoing, because failure of either normative competence or situational control violates the norm that blame and punishment be reserved for those who had a fair opportunity to avoid wrongdoing. If we treat the fair opportunity to avoid wrongdoing as the key to responsibility, we get the following picture of the architecture of responsibility.

The fair opportunity conception of responsibility provides one plausible framework for understanding the retributivist conception of culpability as blameworthiness, and we will draw on it going forward in our understanding of culpability and the limits of responsibility.