

## Rational Capacity as a Condition on Blame

Pamela Hieronymi  
hieronymi@ucla.edu

In “Rational Capacities,” Michael Smith takes up Gary Watson’s puzzle about how to distinguish between cases of recklessness, weakness of will, and compulsion.<sup>1</sup> Watson provides an example:

Suppose that a particular woman intentionally takes a drink. To provide an evaluative context, suppose she ought not to have another because she will then be unfit to fulfill some of her obligations. Preanalytically, most of us would insist on the possibility and significance of the following three descriptions of the case: (1) the reckless or self-indulgent case; (2) the weak case; and (3) the compulsive case. In (1), the woman knows what she is doing but accepts the consequences. Her choice is to get drunk or risk getting drunk. She acts in accordance with her judgment. In (2) the woman knowingly takes the drink contrary to her (conscious) better judgment; the explanation for this lack of self-control is that she is weak-willed. In (3), she knowingly takes the drink contrary to her better judgment, but she is the victim of a compulsive (irresistible) desire to drink. (41–42)

Distinguishing these three categories is important, because they are thought to license different responses to the drinker. While we blame both the reckless drinker and the weak-willed drinker, we do so in different ways, or for different things; the compulsive drinker, on the other hand, is excused in virtue of her compulsion. Although matters of moral importance thus hang on distinguishing these cases, providing a way to distinguish them proves difficult.

Watson believes that both the weak-willed and the compulsive drinker act contrary to their better judgment. This distinguishes them from the reckless drinker. Distinguishing them from one another proves more difficult.

In a natural first attempt, one might claim that the weak-willed drinker could have resisted the temptation while the compelled drinker could not have—she was compelled.

---

<sup>1</sup> See Michael Smith, “Rational Capacities, or: How to Distinguish Recklessness, Weakness, and Compulsion,” *Ethics and the A Priori* (Cambridge: Cambridge University Press, 2004): 114–35; Gary Watson, “Skepticism about Weakness of Will,” *Agency and Answerability* (Oxford: Oxford University Press, 2004): 33–58.

Watson is skeptical, though, about distinguishing weakness of will from compulsion by appeal to whether this particular agent, given her psychology at the time, could have resisted temptation. After all, it is equally natural to think that, given the weakness of her will, the weak-willed drinker could not have resisted the temptation.<sup>2</sup>

Watson instead draws the distinction by appeal to the “capacities and skills of resistance which are generally acquired in the normal course of socialization and practice . . . which we hold one another responsible for acquiring and maintaining” (50). On this account, “weak-willed agents fall short of ‘reasonable and normal’ self-control, whereas compulsive agents are motivated by desires which they could not resist even if they met those standards” (50). Watson’s account, then, draws the distinction between weakness and compulsion by appeal to general ethical standards, standards that are set by what can reasonably be expected of people, given what is normally the case. In certain kinds of cases, we expect people to resist; in others, we do not. People who give into temptation in cases in which we expect people to resist are weak-willed. Those who do not resist in cases in which we do not expect people to resist were compelled.<sup>3</sup>

Watson’s account thus allows that people are sometimes blameworthy for failures that they could not have avoided, at the time: the weak-willed person, who does not have the capacities and skills of resistance we normally expect, may, for that reason, be unable to resist a given temptation. Still, on Watson’s view, we can hold her responsible, so long as she is “constitutionally capable of meeting [normal] standards” of self-control (52). What

---

<sup>2</sup> Watson provides a number of arguments for his skepticism; I have merely gestured at an intuition.

<sup>3</sup> In fact, Watson appeals not directly to *kinds of cases*, as I have done, but to “whether persons of normal self-control would resist *desires of the same strength*” (italics added).

one needs, according to Watson, is an underlying capacity to develop one's capacity for self-control.<sup>4</sup>

In contrast, Smith assumes that, if we are correct in holding a person responsible for something, then there is some more robust sense in which that particular person, in that particular situation, could have done otherwise. He opens his article with his own account of Watson's three cases:

We blame the woman who is reckless or self-indulgent, and what we blame her for is having the wrong belief about what she should do in the circumstance. She believes that the value associated with having another drink makes it worthwhile for her to risk being unable to satisfy some of her obligations, whereas we disagree. . . . [I]t is . . . important that we blame her just to the extent that she could have believed otherwise . . . It would be totally inappropriate if she lacked the capacity to evaluate such evidence, or if, though she possesses this capacity, her belief was the product of (say) self-hatred which she could neither acknowledge nor get rid of. (114–15)

In contrast, the weak-willed drinker has the correct belief. We blame her for her failure to act in accord with that belief. And again, “[b]laming her is appropriate to the extent that she could instead have exercised self-control and desired otherwise” (115). The compulsive drinker, due to her irresistible desire, lacks this ability to have done otherwise, and, according to Smith, this is precisely why we excuse her.

So, according to Smith, we blame the reckless drinker only if she could have believed differently and the weak-willed drinker only if she could have acted differently; we excuse the compulsive drinker because she could not have acted on her correct belief. Smith believes that each of these ‘could’ claims “signify the presence (or absence) of a

---

<sup>4</sup> To be precise, according to Watson one needs not only the underlying capacity to develop the ability to control oneself, if given favorable conditions for its development, but also to have been given such favorable conditions. (I take it this follows from Watson's claim that we hold the weak responsible only if “we believe they could and should have developed and maintained” the ability to control themselves (52).) Thus, while Watson allows that a person can be held responsible for failures that he or she could not have avoided *at the time*, he would presumably deny the stronger claim that one can be held responsible for failures that one could not have avoided because one was not given circumstances favorable for the development of self-control. I will suggest that which claim is true might depend on what is meant by “holding responsible.”

rational capacity which we take to explain the relevant behavior.” His task, in the article, is “to say what, precisely, makes it the case that someone has (or lacks) such a rational capacity” (115). Smith, then, takes up the challenge of rebutting Watson’s skepticism: he hopes to ground the distinction between the reckless, the weak-willed, and the compulsive drinker in facts about what she could have done, given the particulars of her psychology at the time.

#### CONDITIONING BLAME

Smith’s approach is the one many people find natural. Many seem to think that the appropriateness of blame turns on the particular psychological capacities of the agent in question at the time of action. Of course, everyone should agree that the appropriateness of blame turns on the possession of certain very general rational capacities, capacities required even to be an agent or to be the kind of creature who can stand in interpersonal relations with others, such as the capacity to form beliefs, execute intentions, revise one’s beliefs and intentions in response to changing circumstances or opinion, recognize one’s own reasons for thinking or acting in a certain way, and recognize others as conscious creatures who, like oneself, form beliefs and intentions. But many find it natural to think that the appropriateness of blame depends, in addition, on a more specific ability: the agent’s psychology at the time of action must be such that, holding the particulars of that psychology more-or-less fixed, he or she could have responded better in the circumstances in which he or she responded poorly. Blame is appropriate only if the individual blamed had some more specific, local ability to have done otherwise.

It is worth noting, at the start, that this line of thought separates our moral capacities from capacities of other sorts. Typically, our capacities develop as demands are put upon us to exercise them well—beyond our current ability. We usually learn to think more clearly, write more smoothly, run more swiftly, plan more effectively, or sing more beautifully because someone (often oneself) expects better of us. Moreover, in many areas of adult life—in one’s career, in one’s role as teacher or parent, in one’s position as chair or as second tenor—the demands one is under remain insensitive to one’s own particular shortcomings; one’s capacities develop as one tries to meet them. If, in contrast, we cannot be put under genuine moral demand until we are already able to satisfy it, the development and exercise of our moral capacities would be, in this respect, fairly exceptional.

As far as I understand it, the motivation for thinking moral demands exceptional in this way starts with the thought that failing to meet moral demands is a far more serious business than failing to meet other sorts of demands. So it is. Somehow from that fact we are meant to arrive at the view that one cannot be charged with inadequate moral performance or moral failing—one cannot be morally blamed—unless one had the ability to have satisfied that very moral demand. As this thought is sometimes put, ‘ought implies can’—or, more properly, blame implies can.<sup>5</sup>

The doctrine that blame implies can seems to me far less obvious than another, competing doctrine, (which I will call) the doctrine of original sin—the doctrine that

---

<sup>5</sup> I believe this line of thinking can be found, explicitly or implicitly, in much of the important work on the topic, including R. Jay Wallace’s *Responsibility and the Moral Sentiments* (Cambridge: Harvard University Press, 1996), Gary Watson’s “Two Faces of Responsibility,” *Philosophical Topics* 24 (1996): 227–48, and Susan Wolf’s *Freedom within Reason* (New York: Oxford University Press, 1990). Scanlon considers it in “Blame” (unpublished ms.). I have argued against it in “The Force and Fairness of Blame,” *Philosophical Perspectives* 18, no. 1 (2004): 115–48.

people enter the world and (often enough) continue in it while unable to meet the moral or interpersonal demands that are nonetheless legitimately placed upon them. Original sin seems overwhelmingly plausible, given the facts on the ground: Human development is a hazardous affair. People often arrive at adulthood insensitive, self-absorbed, greedy, petty, touchy, and shortsighted. Sharing a world with others requires the opposite dispositions. Some story must be told as to why, exactly, legitimate moral demand should recede in the face of psychological inability.<sup>6</sup>

I grant that two such stories could be told. First, by appealing to the nature of moral demands, one could explain why those demands should lose their grip in the absence of the very general rational abilities mentioned earlier (the ability to form and revise beliefs and intentions, to reflect upon one's own reasons, and recognize others as rational creatures). Moral demands, one might think, are the demands put upon us as members of interpersonal relationships. If a creature lacks the abilities required to stand in such relationships, it does not make sense to put that creature under these demands. Just as it would be absurd to expect the mute to sing on key, so it would be absurd to place moral demands on a creature that could not form intentions or recognize itself and others as (broadly) rational.

Another story could be told about why the particular demands people are placed under—the content of those demands, so to speak—should be limited by general facts

---

<sup>6</sup> “Moral demand” is used to mark those moral expectations the disappointment of which incurs blame or moral criticism. Confusion may result if one takes these demands to be identical to “obligations;” our obligations may be more strict than legitimate moral demands (so understood). It may be possible to blamelessly violate an obligation. By stipulation, it is not possible to blamelessly violate a legitimate moral demand. (What is meant by “blame” is obviously crucial, here. It will be examined below. For now, note that if one thinks of blame simply as “criticizing someone for something he or she could have avoided,” there will be no room for the current dispute; what I have called the doctrine of original sin will contain a contradiction. To avoid begging the question, we should not simply assume this conception of blame.)

about typical human weaknesses. One might think, as Watson seems to, that we should excuse those who are subject to temptations to which anyone, or even most people, would succumb. Just as it would be unreasonable to write a hymn whose notes fell outside the range of the average congregation, so it would be unreasonable to count among ordinary moral demands acts of heroism of which very few are capable.<sup>7</sup>

But the thought that Smith and others find plausible goes beyond these, to claim that the content of reasonable moral demand varies with the particular abilities of a given agent, at a time—they recede in the face of local inability. The reasonably written score would respond to the particular (and perhaps changing) ability of each member of the choir.<sup>8</sup>

Smith does not explain why moral demand should recede in the face of such local inability. He takes it as obvious that it does and tries to articulate in what, exactly, the required abilities consist. I propose to look at his account and ask, of the capacity Smith isolates, why it should condition legitimate blame. I will suggest it does not. I hope that examining the difficulties with Smith's proposal will suggest a broader thesis, viz., that moral demand need not recede in the face of an individual's inability to meet it. Rather, legitimate moral demand requires only that one have the more general capacity to stand in interpersonal relationships—a capacity which may develop into a robust ability to govern oneself morally, as one is put under moral demands.

---

<sup>7</sup> To elaborate: if one thinks of moral demands as those that regulate social life, one might also think they must be sensitive to what human beings are normally like in the same way that hymns—songs to be sung by the entire congregation—should be written with a view to the limitations of the average congregant.

<sup>8</sup> Again, the score is here identified, not with obligation, but with what can be reasonably demanded of a particular person, on pain of moral criticism and blame. At issue is whether it is reasonable to demand that people to live up to standards that they are not currently capable of meeting. In many areas of life, it is. Many think that, when it comes to morality, it is not.

## SMITH'S TREATMENT OF JOHN

Smith approaches his task (of articulating the relevant sense of capacity) by first considering the case of John:

John is in the middle of a complicated philosophical argument with someone when she asks him a crucial question to which he doesn't have an answer. Let's assume that there is an answer, one which supports the line of argument that John has been defending. He thinks the question through carefully, but he doesn't think of the answer at the time. However, later on that night he comes to realize what the answer to the question is. (116)

Smith fills this case out in two different ways. In the first, John leaves the conversation, goes home, and reads some papers on the topic. "While reading through these papers, he learns what the answer . . . is. He has the 'Now I see!' experience . . ." (116). Smith calls this the case of Ignorant John. In the second case,

John doesn't need to read any further material . . . The answer occurs to him while he is driving home . . . Moreover, when it occurs to him he is overwhelmed not by the 'Now I see!' experience, but by the 'Oh dear! Of course!' experience. (117)

Smith calls this the case of Blanking John.

Smith's thought is that Blanking John could, whereas Ignorant John could not, have thought of the answer in the original conversation. He spends time working out the sense of 'could' that makes this thought true.

Smith first dismisses the libertarian's attempt to mark the difference by appeal to whether there is "a possible world which is identical in history and causal laws to the actual world . . . but in which [Blanking John] instead thinks of the right answer" (117). He points out that, given the dialectical situation, we should assume that there is no such world: "if the libertarians were right [that there were such a world] then we would expect that to emerge in what follows" (118), that is, we would expect to be forced to that allow for such a world, to accommodate the relevant capacity. Thus, rather than assume such a

world from the beginning, we should instead think about possible worlds while allowing for small changes in history and laws.

Accordingly, Smith next considers whether we can analyze the sense in which Blanking John could, but Ignorant John could not, have thought of the answer by appealing to the fact that “the possible world in which Ignorant John thinks of a response is much more dissimilar to the actual world than the possible world in which Blanking John thinks of a response” (119).

A variant on Frankfurt’s famous case shows this analysis inadequate. In the variant case, Blanking John blanks because an evil scientist has tampered with his brain. Here, the possible world in which Blanking John thinks of the correct answer is further from the actual world than the world in which Ignorant John does (because eliminating the scientist is more difficult than eliminating the ignorance). Smith suggests that, to avoid the difficulty, we must “zero in on what makes it true that Blanking John has the capacity.”

Accordingly, he next considers the view that

Blanking John could, whereas Ignorant John could not, have thought of the right response . . . because . . . [a]bstracting away from all those properties that could have an effect on what either of them think except the relevant properties of [their] brains, the possible world in which Ignorant John thinks of the right response is much more dissimilar to the way the actual world is than the possible world in which Blanking John thinks of the right response. (122)

Yet even this narrow focus on the brain will not do. It is shown inadequate by the case of Fluky John. As with Blanking John, the answer occurs to Fluky John without his consulting any other sources. However, when the answer occurs to him, “he is . . . overwhelmed . . . by the ‘Why on earth did that occur to me!’ experience” (123). He has “a completely fluky flash” (123). Fluky John is a counter-example to the going proposal,

because the change required in Ignorant John's brain to get to the world in which Ignorant John has such a fluky flash might be smaller than the change required in Blanking John's brain to get to the world in which Blanking John thinks of the answer.

Smith draws the following moral: any attempt to account for the sense in which Blanking John could have thought of the answer while Ignorant John could not which appeals only to "single possibilities" will fail, because "capacities are essentially general or multi-track in nature, and that they therefore manifest themselves not in single possibilities, but in whole rafts of possibilities" (123). In order to have the capacity to have thought of the right answer in the circumstances, Blanking John

must . . . have had the capacity to think of the answer to a whole host of slight variations on the question that he was asked, variations in the manner in which the question was asked, and perhaps in the exact contents of the question, and in the time of the question, and so on. (124)

So Smith arrives at the following analysis of the relevant sense of capacity:

[W]e engage in a triangulation exercise. Begin from the fact that, abstracting away from all of those properties that could have an effect on what any of them thinks except the relevant properties of their brains, in those nearby possible worlds in which Blanking John is asked a whole host of similar questions, he has the right thought in response. . . . whereas in the nearby possible worlds in which both Ignorant and Fluky John are asked that whole host of similar questions, they systematically fail to have the right thought in response. From this we draw the conclusion that, in that same nearby region of logical space, Blanking John is, whereas both Ignorant John and Fluky John are not, having the right thought in response to the question he was in fact asked at the time it was asked. This, accordingly, is why the possible world in which Blanking John has the right thought. . . . is nearer to actuality than is the possible world in which Ignorant John and Fluky John have the right thought. . . . It is why Blanking John could, whereas Ignorant John and Fluky John could not, have thought of the right response to the question when asked. (124)

To this account, Smith adds a further point: If Blanking John does indeed have a rational capacity, his correct responses in the nearby possible worlds must be explained by some underlying structure. That is, it must not be the case that

each individual answer he gave to a question he was asked was the result of some aspect of his internal condition that was dedicated to giving exactly that answer in response to exactly that formulation of the question, an aspect which has nothing in common with any other aspect of his internal condition. (125)

With that caveat in place, Smith ends his analysis of the sense in which Blanking John could, whereas Ignorant John could not, have thought of the right answer. Blanking John is in possession of what Smith calls a rational capacity, as shown by the truth of a raft of counterfactuals when those are explained by an appropriate underlying structure.

#### SMITH'S TREATMENT OF THE DRINKER

This might be an entirely satisfactory analysis of the difference between the Johns. The crucial points seem important and correct. Smith then applies this analysis to Watson's case, in an attempt to distinguish between the reckless, weak, and compulsive drinker. I will first explain the application, and then ask why this sort of capacity is required for blame.

I will focus on Smith's treatment of the reckless drinker, as he applies his analysis most straightforwardly in this case. Smith believes that we blame the reckless drinker for having the wrong belief, and, further, that we do so licitly only if she could have believed differently.<sup>9</sup> Recall Smith's condition on blame: "It would be totally inappropriate [to blame her for this belief] if she lacked the capacity to evaluate [the available] evidence or if, though she possesses this capacity, her belief was the product of (say) self-hatred which she could neither acknowledge nor get rid of."

Smith determines whether the reckless drinker has the capacity to form the better belief by employing the apparatus he developed to account for the Johns:

---

<sup>9</sup> I doubt that we blame the reckless drinker for having the wrong belief; we blame her for drinking recklessly. (To my mind, this is also what we blame the weak-willed drinker for doing. The difference is simply that the weak-willed drinker acts against her better judgment, and that fact is mitigating.) Nonetheless, we can, I believe, blame people for having certain beliefs, and the belief Smith picks out could be among them. So I will accept Smith's analysis, for the sake of argument.

We must determine whether or not [the] woman who fails to form the belief [that the value of another drink does not make it worthwhile to risk being unable to fulfill her obligations] has the capacity to form [that] belief. . . . We must therefore abstract away from all those properties that could have an effect on what she believes except the relevant properties of her brain, and we must then ask whether a whole raft of counterfactuals is true of her. Would she have formed a whole host of similar beliefs in response to similar evidence? If she would have, then, the suggestion is, assuming that there is a relevant structure which underlies the truth of those counter-factuals, we can triangulate to the conclusion that, in that same nearby region of logical space, that same woman forms the right belief in response to the evidence she in fact considers. If all this is so, then it follows that she has the capacity to form the right belief in response to the evidence she in fact considers. (126)

We thus have an account of what it would be for the reckless drinker to have the capacity to evaluate the evidence. However, to secure blame, we must also rule out the second disjunct: her faulty belief must be explained by her failure to exercise this capacity, rather than by something like self-hatred. To rule out the latter,

we must ask [again] whether the woman had the capacity to form the correct belief . . . , but now instead of abstracting away from her various emotions, we take her emotions into account. If, for example, she does suffer from self-hatred, does her self-hatred undermine the truth of the relevant counter-factuals? If so then we conclude that her self-hatred was the explanation, rather than her failure to exercise her capacity. (126–7)

If her self-hatred (or something similar) was the explanation for her belief, rather than her failure to exercise her capacity, then, according to Smith, we do not licitly blame her.

Whether blame is appropriate turns on what explains her failure to form the right belief.

Importantly, her failure to exercise her capacity might, according to Smith, receive a further explanation. Perhaps, if she had considered certain past experiences, or engaged in certain imaginative exercises, she would have formed the correct belief. If so, then it was her failure to do *these* things that explains her failure to exercise her capacity to have formed the right belief. Importantly, though, according to Smith the explanation of any *culpable* failure will end at a failure to exercise some or another capacity. Thus, while we might be able to give an intermediate explanation of why the reckless drinker failed to form the right belief—she failed to consider her past experiences, or failed to engage in

certain imaginative exercises—given that she is culpable for her failure, her failure to do these things must, in turn, be explained by her failure to exercise her capacity to have done so (her capacity to have considered her past experiences, or to have engaged in imaginative exercises). Smith concludes that

at some point we will have to rest content with explaining the reckless woman's failure to exercise her capacity . . . by saying that she simply blanked, and for that we will be unable to give anything by way of explanation—not, at any rate, without retreating to brain science. (127–8)

One might have misgivings about holding people responsible for failures for which there is no explanation, short of brain science. Such a failure might have been exactly the sort of thing one would have wanted to excuse—a phenomena for which the only explanation is some bodily event which cannot be rendered in psychological terms. But Smith does not see a problem, here. He says,

Importantly, however, this will give us no reason whatsoever to suppose that we cannot hold the reckless woman responsible for failing to think of whatever it is, and hence for blanking. For so long as we are right to assume that her failure of thought, or her blanking, occurs in a suitable context of nearby possible worlds in which she does have that thought, it thereby follows, analytically, both that she could have had the thought or that she could have not blanked, in the sense of having had the capacity to have the thought, or not to have blanked, and that her failure to exercise her capacity is the relevant explanation of her failure to have the thought, or of her blanking. And this, in turn, is what legitimizes our holding her responsible. (128)

One might doubt whether that which follows from the truth of Smith's counterfactuals is what legitimizes our holding the woman responsible. The next section considers whether or how the sort of capacity that Smith has identified should be relevant for moral responsibility.

#### SMITH'S SORT OF CAPACITY AND MORAL RESPONSIBILITY

Start with the fact that we do not blame Blanking John. Why not? He was, after all, asked a question, a question he had the capacity to answer, and he failed to do so. The

relevant explanation of his failure is the fact that he failed to exercise his capacity. Yet that fact does not seem to legitimate blame.

A charitable reading of Smith would insist that failing to exercise a capacity is a necessary but not a sufficient condition for legitimate blame: if someone is legitimately blamed for some failure, that failure must be or be explained by the failure of some capacity, but not every failure of a capacity is blameworthy. We should wonder, though, what makes the difference. Why is the reckless drinker blamed for failing to form the right belief while Blanking John is not?

One obvious difference between Blanking John and the reckless drinker is the topic on which they fail to form the right belief. Blanking John fails to form a belief about some complex philosophical argument. The reckless woman forms the wrong belief about the value of taking another drink in relation to the importance of her obligations. The content of her belief concerns the worth of other people—the importance of the obligations that they place upon her, relative to her own desires or pleasures. Someone who believes that the value of having another drink outweighs the importance of one's obligations has an offensive belief—a belief whose content portrays an overvaluation of oneself or one's own pleasures as against the needs or claims of others. One might think that holding an offensive belief is itself offensive. It is a moral failing. Smith would add: so long as one could have believed differently.

There is a minimal sense in which to call something a belief is already to guarantee that one could have believed differently. To call something a belief, rather than a fixation or illusion or stubbornly recurrent thought, is to assign it a particular kind of place in a

person's mental economy. It represents the person's take on what is the case, and it must interact more or less rationally (though far from perfectly rationally) with the person's other beliefs and attitudes. So, if the person changes her mind sufficiently, her belief will change. A minimal degree of rational capacity—the capacity to have beliefs—will secure this much control and possibility.<sup>10</sup>

But, of course, Smith thinks a more robust sort of capacity to have believed differently is required before we hold the reckless drinker responsible for her offensive belief. The relevant raft of counterfactuals must be true of her: in a host of similar worlds, she must have formed an inoffensive belief in response to similar evidence. She must be like Blanking John rather than Ignorant John.

Why think someone with an offensive belief needs to have a capacity like Blanking John's, before blame is licit?

Sometimes conditions on the legitimacy of blame are supported by the thought that blame is a kind of negative treatment, a kind of punishing or sanctioning, or perhaps something like chastising or guilt-tripping. It seems plausible that these activities are legitimate only if their target had an adequate opportunity to avoid the negative treatment. It seems unfair to punish someone who did not, at some point, have a relatively obvious opportunity to have avoided the punishment.<sup>11</sup> Often those who take Smith's general

---

<sup>10</sup> Smith himself seems to rely on something like this thought in accounting for the weak-willed drinker. Though she has the right belief, she does not act upon it. She could have acted upon it, though, because, as a rational person, she has the capacity to have coherent desires.

<sup>11</sup> If one thinks of blame as something akin to the final judgment, the passing out of divine reward or punishment on the basis of one's moral record, then one might well think it should be conditioned by the opportunity to have done otherwise. Famously, the doctrine of original sin sits rather uneasily with the final judgment (seeming—paradoxically—to require grace).

approach think along these lines: blame is a kind of negative treatment, and before we engage in it, we must consider the particular psychology of its target, in order to determine whether, given the facts of that psychology, he or she confronted a genuine opportunity to have avoided the blame.

Notice, though, how odd it would be to apply Smith's analysis to this end: Smith has shown that Blanking John had the capacity to have answered correctly, in the sense that, in a host of nearby possible worlds, he would have answered correctly. So Blanking John *could* have answered correctly, one might object, only in the sense that Blanking John *might* have answered correctly. Smith has secured a metaphysical possibility, here—in fact, a strong one—but he has not thereby secured any sense in which it is *up to John* whether he thinks of the right answer, when asked. Possibility—even strong possibility—does not amount to control.

Suppose I have a heart attack. It may well be that, in a host of similar possible worlds, I do not have a heart attack. Further, the fact that I do not have a heart attack in those worlds may be explained by the underlying structure of my cardio-vascular system. Thus, I have the capacity, in Smith's sense, to have not suffered the heart attack. The truth of this claim does nothing to show that it was up to me whether I had a heart attack; neither does it show that I had an opportunity to avoid the heart attack.

Things are worse for Blanking John. Recall that, according to Smith, if Blanking John is culpable for his blanking, the explanation of his failure must run out at the fact that he simply failed to exercise some capacity. His failure to exercise the relevant capacity is itself inexplicable. But anything that is both inexplicable and, given the

circumstances, improbable (as shown by the raft of counterfactuals) will also be uncontrollable, and therefore unavoidable.<sup>12</sup>

Thus, if one is thinking of blame as a kind of punishment or sanction, it can seem cruel to blame those who are, in the relevant way, like Blanking John, exactly because there is a very salient sense in which they could not have avoided failure: Although John has the capacity to not fail, in the sense explicated, he does not have any means by which to avoid failure. Although he failed to form the correct belief, he had all the capacities required for doing so. For reasons that remain opaque both to him and to us (for no reason short of brain science), these capacities simply did not engage in the circumstance. He just blanked. Sometimes that happens to people.

If one thinks of blame as a kind of negative treatment whose legitimacy requires that its target have had an adequate opportunity to avoid it, one will want an account of how a given response is up to an agent or under his control. Smith's account of capacity does not secure this (in fact seems to rule it out) and so is unlikely to satisfy those who think of blaming as a kind of negative treatment, punishing, or sanctioning.<sup>13</sup>

---

<sup>12</sup> In contrast, my heart attack, though improbable, may be explained by facts over which I might have had some control. Thus, I may have both the capacity, in Smith's sense, not to have a heart attack, and means by which to avoid it. (To be precise, the blanking is, according to Smith, inexplicable short of brain science. So it will not be controllable, short of brain science. But surely the kind of control we are meant to have over our beliefs and actions before we are responsible for them is not the kind secured by knowledge of brain science.)

<sup>13</sup> Presumably, this is the point at which the libertarian will insist that his answer emerges: it is not enough to secure a host of nearby possible worlds (with changes in the past and the laws) in which John thinks of the answer; we must rather hold fixed the past and the laws and look for a possible world in which John thinks of the answer. Such a possible world, the libertarian thinks, is required for John to control what happens in the future. (This last claim is disputable. See, e.g., John Martin Fischer and Mark Ravizza, *Responsibility and Control* (Cambridge: Cambridge University Press, 1998).)

One might have something else in mind, when one thinks of blame. One might, instead, mean something like acknowledging that a person's actions, attitudes, decisions, or reasoning do not take other people or their interests properly into account. P. F. Strawson points out the importance we attach to what he calls "the quality of other's wills towards us" and towards others.<sup>14</sup> He characterizes certain attitudes, such as resentment, indignation, and gratitude, as reactions to (our perceptions of) the quality of other's wills. If someone recognizes that your actions or attitudes betray some lack of concern or respect, that person might experience resentment or indignation. Further, he or she might mistrust you, withdraw esteem, or retreat from intimacy. In general, if others recognize that you have not taken their interests rightly into account, this recognition has certain relatively predictable effects on the web of your relations with them—effects which are not straightforwardly voluntary, and which therefore are not well understood as forms of treatment or sanction.

Suppose, then, that this is what we have in mind, when we think of being blamed: having the fact that you did not take others properly into account acknowledged by others, who react accordingly.<sup>15</sup> Blame, in this sense, will obviously require those capacities required to stand in interpersonal relations. But why must a person must be like Blanking John, rather than Ignorant John, before such blame is legitimate? One might think this the opposite of the truth.

---

<sup>14</sup> P. F. Strawson, "Freedom and Resentment," *Proceedings of the British Academy*, 48 (1962): 5–20, citations as reprinted in Gary Watson, ed. *Free Will* (Oxford: Oxford University Press, 1982): 59–80.

<sup>15</sup> So understood, blame can come in various flavors and strengths. We can acknowledge, even insist, that certain sorts of blaming reactions are inappropriate, in light of this or that circumstance, while also insisting that blame, in this wider sense, is appropriate.

If the reckless drinker's offensive belief is like Blanking John's lack of belief, then in nearby possible worlds in which she encountered similar circumstances, she formed the right belief. Her "recklessness," then, is not very stable or deep. It is, itself, a little fluky. One might think that, to the extent that a belief is in this way fluky, to that extent it need not inform one's interactions with the woman in any serious or lasting way. If we learned that a given offensive belief was a slip, due to a blanking, which, in a host of nearby possible worlds, would not have happened, that fact would be mitigating.

In contrast, to the extent that the reckless drinker is like Ignorant John, her offensive belief is relatively stable. She will form the same sort of belief about the relative importance of her own pleasures or desires and other's interests, in a host of similar circumstances. She would need some new experiences—alogous to reading the articles—before she changes her mind. This fact, it seems, would alter one's relations to her quite substantially.

Suppose, then, that the reckless drinker's offensive belief is to be explained by the "self-hatred she is unable to acknowledge or get rid of." Presumably, the inability here is of the same sort: in nearby possible worlds she does not acknowledge or get rid of it. Why should such a condition undermine blame? One might think the fact that an offensive belief is supported by a relatively stable self-hatred (or self-absorption, or inability to sustain a focus on the interests of others) would inform, rather than suspend or delegitimate, the effect of her belief on one's relations with her.

Perhaps Smith means to suggest that, if the failure is explained by incorrigible self-hatred (or insensitivity, or insecurity . . . ), we should withdraw from the sorts of

relationships of which blame is a part. Strawson, famously, allows that we can abandon our concern with the quality of others' wills and instead adopt what he calls an "objective attitude" towards them. We then see someone

as an object of social policy; as a subject for what, in a wide range of sense, might be called treatment; as something certainly to be taken account, perhaps precautionary account, of; to be managed or handled or cured or trained... (66)

According to Strawson, we adopt the objective attitude when certain facts about the person (e.g., insanity or "particularly unfortunate formative circumstances") convince us that he or she is not a participant in normal adult relationships, not "a term of moral relationships" or "a member of the moral community" (73). We can also recourse to his stance for other reasons, e.g., "as a refuge from the strains of involvement" (67), or even no reason.

So perhaps Smith means to suggest that, unless a person has the sort of capacity he has identified, we should not concern ourselves with the quality of that person's will but should instead adopt these more objective attitudes. But one might think this would be giving up too early, expecting too little in the way of moral development or improvement, retreating too quickly from the strains of involvement.

On neither of these two understandings of blame does the sort of capacity Smith isolates seem a natural condition. Perhaps a third possibility is being overlooked. Short of one, it remains unclear exactly why the kind of capacity Smith articulates so clearly in the case

of Blanking and Ignorant John should be the kind of capacity required for legitimate blame.<sup>16</sup>

At this point, we have simply retraced certain familiar lines of debate. If one thinks of blame as a kind of punishment, one might think it legitimate only if its target had an adequate opportunity to avoid it. Someone in this frame of mind will not find Smith's analysis satisfactory, because, although Smith has secured a sense in which it is possible that the person did otherwise, the person does not have the right sort of control over whether that possibility obtains.

People who think of blame as a kind of punishment sometimes also think that, when assessing whether someone had a fair opportunity to avoid it, we should hold that person's psychology, at the time, more-or-less fixed, and ask whether, given the particulars of that psychology, the person could have done otherwise. Answering this question affirmatively can seem to require libertarian freedom: while holding fixed the present psychology, we must somehow allow the person to control which of various futures obtains.<sup>17</sup>

If, instead, one thinks of blame as the effect on one's relationships of the acknowledgment by others that one's actions and attitudes have not taken other people or

---

<sup>16</sup> Again, one might mean, by "blaming," "chastising (or punishing, or thinking badly of) someone on account of something that person might have avoided," where the "might" is custom-fit to Smith's sense of capacity. Smith's sense of capacity would then obviously be required for it. Perhaps blaming, so understood, is a recognizable practice; it is not one I would defend.

<sup>17</sup> One need not move from the thought that the fairness of a sanction requires an adequate opportunity to avoid it to the claim that the fairness of a sanction requires libertarian freedom. T. M. Scanlon, e.g., separates blame from punishment. He allows blame to track the quality of a person's will in roughly the way sketched above. Punishment, in contrast, he sees as a means of protecting certain important societal goods (such as public safety). His approach yields an entirely mundane, political (not metaphysical) account of what constitutes a "reasonable opportunity to avoid." See "Responsibility," in his *What We Owe to Each Other* (Cambridge, MA: Harvard University Press, 1998).

their interests properly into account, then blame seems grounded in the facts of the actual world. If blaming a person is simply the adjustment of one's attitudes and relations in reaction to the quality of another person's will, it is quite unclear why blame would require that, in nearby possible worlds, the person would have had a better will. In fact, it might seem that, to the extent that Smith's raft of counterfactuals is true, or to the extent that the failure is brute or inexplicable, blame, in this sense, will be mitigated. To the extent that the quality of a person's will would vary in nearby possibilities, her current ill will might seem less important to one's on-going relations with her.

In fact, if one thinks of blame in this second way, it is unclear why legitimate moral demand should recede in the face of a particular individual's inability to meet it. Rather, vulnerability to legitimate blame might seem secured by the possession of the more general rational and interpersonal abilities referenced earlier: the ability to form and revise beliefs and intentions, to recognize one's own reasons for acting, and to recognize others as conscious, rational creatures. Given these more general abilities, one will inevitably be in the business of forming beliefs and other attitudes concerning the importance of other people and their interests relative to oneself and one's own interests, and acting accordingly. Others can legitimately demand that they and their interests be acknowledged, in one's actions and attitudes, on pain of altering their relations with you in certain characteristic ways. Of course, exactly what they can legitimately demand—the content of their legitimate demands—will be sensitive, in certain ways, to the particularities of the case—perhaps in the sort of way that Watson suggests. But legitimate moral demand need not recede in the face of a particular individual's inability

to meet it. Suppose that one's personal development is such that one does not currently have the ability to recognize others' interests (due to insensitivity, self-absorption, or self-hatred) until they bring it vividly to one's attention; suppose, further, that one is so constituted that one cannot, even then, bring oneself to properly respect those interests in one's actions. Moral inability of this sort need not undermine, but might rather ground, such reactions as resentment, indignation, distrust, withdrawal of esteem, and refusal of intimacy. Thus, such inability need not make the others' demand—that they and their interests be acknowledged—any less legitimate.

On the proposed picture, the possession of certain very general, broadly rational, capacities lands one in interpersonal relations. Once in such relations, one's activities will trigger a web of non-voluntary but legitimate interpersonal reactions from others, should one mistake their value. These reactions might, in turn, prompt the development of one's own moral abilities. But, of course, the facts of one's existence might be such that, tragically, one cannot thereby develop. One might be unable to transcend one's own self-absorption, self-hatred, defensiveness, or pride in order to leverage the reactions of others for the purpose of one's own improvement. If one is so unlucky that, in all nearby possible worlds, one cannot overcome one's moral shortcomings, that fact seems to guarantee that one will, in all those worlds, likewise fail to enjoy certain sorts of positive interpersonal relationships. Instead, one's relations will be of another character—one will be blamed (resented, distrusted) or, eventually, written-off, avoided, or subjected to

treatment. Such, it seems, are the hazards of original sin—of entering the actual world in need of moral and personal development, while sharing it with others equally real.<sup>18</sup>

---

<sup>18</sup> “Others equally real” is, of course, Thomas Nagel’s apt turn of phrase. I owe thanks to many for helpful comments or conversation, including Barbara Herman, Sean Kelsey, John Martin Fischer, and Gary Watson. Special thanks are due to Michael Smith, both for providing the original article and for his gracious help in conservation.