

# THE SMART THEORY OF MORAL RESPONSIBILITY AND DESERT

Richard J. Arneson

Published in *Desert and Justice*, ed. Serena Olsaretti (Oxford: Oxford University Press, 2003)

The "Smart" of my title is J. J. C. Smart. He has proposed an austere version of compatibilism.<sup>1</sup> The generic doctrine of compatibilism holds that the claim--that all human choices are events in the physical world that are caused either deterministically or indeterministically--is compatible with moral responsibility and desert.<sup>2</sup> According to Smart's version, one is morally responsible for a choice one makes just in case praising or blaming, rewarding or punishing one for making the choice would produce good consequences by altering the future behavior of oneself or others. Compatibilism of this ilk does not include the assertion that free will and the causation of choices are compatible, and indeed Smart repudiates the libertarian idea of free will on the ground that it is logically incoherent and does not consider whether some watered-down notion of free will might make sense. If compatibilism plus determinism equals soft determinism, Smart's doctrine merits the label "hard soft determinism."

This is the position everyone loves to hate. R. Jay Wallace calls it the "economy of threats" approach to the understanding of moral responsibility. If we consider the case in which one holds oneself responsible for one's own wrongdoing, Wallace urges that we should agree that "the economy of threats account of the reflexive case has an almost comically external aspect."<sup>3</sup> Saul Smilansky has given the economy of threats approach yet another label, "effect compatibilism," but echoes Wallace's negative assessment of it. Smilansky writes, "Effect compatibilism is morally grotesque, contrary to any

conceivable set of reflective common moral beliefs, and in light of the case for [Smilansky's alternative account], morally wrong in comparison with the defensible practical alternative."<sup>4</sup> In the same spirit, but more calmly, T. M. Scanlon observes, "The usefulness of administering praise or blame depends on too many factors other than the nature of the act in question for there ever to be a good fit between the idea of influenceability and the idea of responsibility which we now employ."<sup>5</sup>

These objections for the most part issue from a "soft soft determinist" or soft compatibilist perspective (as contrasted with Smart's hard compatibilism).<sup>6</sup> Soft compatibilism asserts that people can be genuinely morally responsible for their conduct in the sense of being truly praiseworthy or blameworthy, deserving of reward or punishment, good fortune or bad fortune, in a world in which all human choices are events caused by prior events. The question arises, what rises and falls with acceptance or rejection of full-blown moral responsibility and desert of the sort soft compatibilism claims to deliver. This essay attends to this question with a view to clarifying the interaction between debates about free will and moral responsibility and debates concerning the role of personal responsibility and desert in theories of social justice.

### **1. The Influenceability Account of Responsibility**

In "Free-will, Praise, and Blame," Smart rejects the idea of libertarian free will on the ground that it is logically incoherent. He proposes that the idea of moral responsibility can be salvaged from the wreckage of free will notions. What he picks up from this junk heap is the thought that ascriptions of responsibility make sense as devices to alter the causation of future acts so that they turn out to be more rather than less desirable. Holding an individual responsible for the quality of her acts is resolving to

impose good treatment on her if the quality is good and bad if the quality is bad. Doing this is ill-advised if the person is not responsive to incentives so that presenting her with blandishments and threats is a waste of time and resources devoted to fulfilling these commitments are resources poured down the drain. Human nature being what it is, the practice of moral responsibility is probably often well advised.

Smart is aware that imputations of blame and ascriptions of moral merit often presuppose that individuals can be responsible for their conduct in a deeper sense that implies the idea of free will that he rejects or something close to it. The task then is to make sure that the recommended notion of moral responsibility is laundered thoroughly so that no trace of this nonsensical intellectual lint remains clinging to it. In this spirit Smart considers praising and blaming. Praise in one sense clearly carries no inchoate suggestion that the object of praise has free will. After all, one may praise a car for its speed and fuel economy and a desert landscape for its stark beauty without in any way hinting at denying that the car and the desert have become what they are purely through the operation of sufficient prior causes (or some mix of causation and chance). To praise is to indicate that the target of praise scores high according to some standard of merit. Smart stipulates that for his purposes praise in this metaphysically innocent sense is to be contrasted with dispraise. Praising is grading high and dispraising is grading low. What then of praise as contrasted with blame? No doubt many who apply these words to significant human actions are consorting with supernaturalism and antiscientific metaphysics. Smart recommends that “a clear headed” person will use the words “praise” and “blame” in such a way that praising is grading high plus holding responsible and blaming is grading low plus holding responsible.

If you declare that Smith, for example, is blameworthy for harassing the cat, but you are brought up short and inclined to retract your declaration on being informed that Smith's harassing conduct is entirely the causal product of forces of heredity plus environment, this is a sign that you are deploying a notion of blame that is metaphysically tinged and unsustainable. You are in the grip of the idea that if, but only if, an individual genuinely has a free choice between good and evil so that the totality of prior causes leaves it open and undecided which way she will choose, then she is responsible for her choice in a deep way such that she deserves to suffer evil if her choice is evil and good if her choice is good quite independently of whether such punishment and reward or threats and promises of such punishment and reward will do anything at all to bring about better states of affairs. According to Smart it would not be a useful exercise to ponder whether or not this conditional is true because the antecedent is definitely false. Moreover, the "only if" part of the statement is just obviously correct. Nothing but libertarian free will could provide a rational basis for deep responsibility. As to whether libertarian free will itself could do the trick, we need not decide, since according to Smart free will is impossible.

The metaphysical element in the notion of responsibility that gets expressed in problematic praising and blaming is what Smart labels "judging." He has in mind judging that the saint truly deserves reward and the sinner truly deserves punishment not as means to a further good end but as desirable in themselves. Judging is what God does when he decides that one person deserves eternal heaven and another deserves eternal hell. All such exercises in discrimination in effect presuppose that when a person behaves badly by moral standards a yellow stain appears on her soul. We may continue

to engage in the discriminations while taking the talk of yellow stains on souls to be metaphorical, but according to Smart there is no sensible way of cashing out this sort of metaphor.<sup>7</sup> Smart concludes, “The upshot of this discussion is that we should be quite as ready to *grade* a person for his moral qualities as for his non-moral qualities, but we should stop *judging* him.”<sup>8</sup>

## **2. Extrapolating from Responsibility to Desert**

Smart's topic is responsibility and determinism, and my primary aim in this essay is to assess his contribution to that topic. In this section I consider to what extent the hard compatibilist line on responsibility, if it proves acceptable, should tend to deflate the significance of desert and of giving people what they deserve. The discussion of this section is continued in sections 8 and 9 of this essay.

Judgments of desert in one range of cases involve nothing more than grading people's performances according to criteria of merit. If ten painters including me reside in my neighborhood, and my paintings are tenth-best, then I deserve to be rated the tenth-best painter in my neighborhood. If my paintings are excellent on an absolute scale of merit, then they deserve to be rated excellent. If I play better tennis than my opponent in a particular match, then I deserve to win this match. These desert judgments do not conflict with anything Smart's position entails.

Judgments of desert can contrast with judgments of entitlement by qualification for prizes and awards. If I very nearly completed a book in a given year that was far more brilliant than any other book published that year, but failed actually to complete the book by some fluke, I may deserve the prize for best-book-of-the-year even though I am not entitled to the prize. Also, one may be entitled to the prize, because the prize

committee has deemed one's book the best in a given year, but not deserve it, perhaps because the committee made a mistake by overlooking a more meritorious entry, or perhaps because the competition in that particular year was subpar.

Desert is sometimes proposed as a fundamental principle of morality or justice: The good things in life should accrue to people in proportion to their moral desert.<sup>9</sup> This principle can be regarded as a regulatory ideal--a standard for designing, assessing, and reforming institutions, laws, and social practices. It can take either a comparative or a noncomparative form. The noncomparative version of the principle holds that a person who is morally virtuous to a specified degree should get cakes and ale and the other good things of life in an amount that is fitting for that particular virtue level. The comparative version holds that a person who is more morally more virtuous than others should get proportionately more cakes and ale and the other good things of life than the less virtuous others. Any position that holds that the quality of people's lives should correspond to their level of moral desert relies on the presupposition that people are genuinely responsible for their voluntary choices and actions. Smart's view is that this presupposition to be sensible must include a belief in free will, which should be repudiated.

One might hold that desert matters without holding that the desert that matters is moral desert or virtue. David Miller takes this line.<sup>10</sup> He asserts that one principle of justice is that individuals should get what they deserve, where what one deserves is not entirely fixed by the rules of existing practices and institutions but sets a standard to which they should conform. In general, a person comes to be deserving of some benefit by acting in a way that the person's society reasonably regards as valuable or admirable.

For example, those who are more economically productive than others produce goods or services for which people are willing to pay more deserve greater economic reward or remuneration.

This type of view interprets desert as compatible with determinism. In a deterministic world one individual may act in a way that her society reasonably regards as valuable or admirable, given any ordinary understanding of what it is to act. The desert theorist of this stripe need not assert anything that Smart denies.

But from Smart's perspective, a worry about desert so interpreted quickly emerges into view. If a desert principle says merely that one whose performance meets a norm (or fails to meet it) should be recognized as such, then desert implies nothing about the proper distribution of benefits and harms to people. But if the principle does have implications for how people should be treated, whether their lives should be made good or bad, what warrants these claimed implications? Why should the quality of my performance, for which I am not claimed to be truly responsible, render it intrinsically morally desirable to heap benefits on me? If the fact that my performance is low-quality rather than high-quality does not lie within my power to control, why think it intrinsically fitting that the high-quality performers should get more of the goods of life than I get? The best response from Smart's perspective is that the idea that one person deserves a better fate in virtue of the quality of her performances should be dropped. The concept of desert might be redefined. One might stipulate that to be deserving is to perform in such a way that rewarding or remunerating one's performance would produce good consequences. Such a stipulation flatly opposes the ordinary notion of being deserving.

### 3. Criticism of Smart on Libertarianism

Return now to Smart's view on responsibility. Despite my admiration for Smart's approach to the topic and indeed his way of doing philosophy, I have to acknowledge that the corners of his analysis don't fit together snugly.

In the course of reviewing J. L. Austin's arguments to the effect that ordinary language is inconsistent with determinism, Smart notes that this issue is not in the end very important because "ordinary language may well enshrine a falsehood."<sup>11</sup> Yet if we turn to Smart's own rebuttal of libertarian free will we find that it consists in an argument to the effect that it conflicts with our ordinary understanding of the terms "determined" and "by chance." Smart's claim is that we might be prepared to revise our understanding of these terms so that what we used to count as an event that occurs by chance we now count as an event that is causally determined or vice-versa. But what does not seem revisable is the idea that these two notions of chance and determination cover all of the logical possibilities. Hence any description of libertarian free will, which must purport to describe events that come about neither by chance nor by causal determination, must be strictly incoherent.

This argument should not persuade the libertarian to give up her cause. So far as I can see, Smart is just pointing to a conflict between the inchoate idea of libertarian free will and our ordinary concepts of chance and necessity. But suppose that when the libertarian notion is articulated, it does clearly conflict with this pair of concepts. If we have been given good reasons to accept libertarianism we will eo ipso have been given good reasons to revise these opposed concepts or junk them and find close substitutes that



do not present the appearance of contradiction that troubles Smart. Here also, ordinary language may embody a mistake.

My quick dismissal might prompt the suspicion that I am being unfair to Smart's argument, which clearly relies on the registration of conflict between our scientific understanding of the world and the medieval notion of free will. But Smart actually cites two separate reasons for rejecting libertarianism, one being its conflict with modern biology and psychology, the other being the purely philosophical difficulty that the free will idea does not admit of coherent formulation. He develops only the purely philosophical difficulty though he acknowledges that the other difficulty is at least equally a strong objection. The purely philosophical difficulty emerges only against the background of our current stock of concepts.

If Smart were correct about the purely philosophical difficulty, the idea of free will is like the idea of a round square, a logical impossibility. But this seems wrong. Free will looks to be a vague and woolly but still recognizably empirical claim. It conflicts with the causal thesis, the claim that human choices are included in the class of events and all events are either fully causally determined according to physical laws or brought about by chance according to physical laws. The causal thesis looks to be true, given current science and its trajectory, but the current state of science does not rule out further unexpected developments in biology and psychology that would undermine it. It would be idle to speculate about what shape such developments might take, that would have the effect of somehow vindicating libertarian free will, and I myself would not bet on the prospect of future science veering off on such a course. But the claim that free will is logically incoherent should not impress us any more than a logical demonstration

by Kant that space must be Euclidean. Hence I find Smart's dismissal of the libertarian free will option to be too sweepingly decisive and prematurely confident.

#### **4. Criticism of Smart on Praising and Blaming**

The idea of moral responsibility in Smart's view comes down to the idea of holding someone responsible. This latter involves the issuing of promises of reward and threats of punishment and the fulfillment of these commitments, all done with a view to bringing about more desirable choices and actions in the future. Smart dissociates the rational core of the practice of responsibility from judging people in ways that, as he thinks, presuppose (some vestige of) libertarian free will. But he does not consider that judging and blaming and shaming in the ways he rejects might be valuable instrumental additions to the practice of responsibility.<sup>12</sup>

Compare two scenarios. In one, I hold myself responsible for being nice to my cat in the sense that I bring it about that I will suffer a penalty if I fail in this regard. In another scenario, in addition to setting up the penalty, I become disposed to grow angry and indignant at myself if I fail to be nice to the cat. I become disposed to flagellate myself with self-blame. Failing to carry out what I regard as my duties to the household pet, I will come to think of myself as a not nice person, one whose unkind character needs reformation, and so on.

Suppose it turns out that judicious application of the proper dose of self-blame makes the practice of holding myself responsible more productive of good consequences overall than would a more austere and mechanical assignment of rewards and penalties minus all the self-assessment and self-reproach. It is a separate and distinct question whether holding others responsible in a way that includes blaming them for misdeeds and

praising them in the judging style for good deeds also can be done in a way that yields better consequences than would be obtainable if moral responsibility were freed altogether from judgment. Suppose the answer here also turns out to be positive. These available gains would then generate a reason to enrich moral responsibility by adding to the practice the blaming and judging that Smart urges we should eschew. The reason is pragmatic. We might choose between a stripped-down conception of moral responsibility (to be deployed when we hold people responsible) and a notion that adds fins and bells and whistles that make it resemble more closely our ordinary idea of what moral responsibility involves by determining which notion of responsibility would yield expectably better results.<sup>13</sup>

This difference between a thinner and thicker notion of responsibility needs clarification. For Smart, attributions of praiseworthiness and blameworthiness make sense only if understood in this way: One is praiseworthy for what one has done just in case the performance scores high according to some appropriate standard and praising or rewarding the performance would influence future performance in some desirable way (more broadly, would produce good consequences). The condition for blameworthiness is analogous. In a thicker sense, one is genuinely praiseworthy just in case these conditions are met along with a further condition. The further condition is that responding to the performance with expression of a positive reactive attitude would influence future performance in some desirable way (more broadly, would produce good consequences). A reactive attitude toward behavior is approval or disapproval of it based on its quality, the behavior being regarded as if chosen with free will. This is what Smart

calls judging behavior, and it is contaminated with the notion of free will that according to Smart makes no sense.

Now compare a puppy's displaying unwanted behavior such as wetting the carpet with a person's behaving in a way that provokes indignation. It might be that the puppy's behavior is blameworthy in the thin sense, whereas the person's behavior satisfies the conditions for thick blameworthiness. If this is true in general, persons tend to be capable of a form of responsibility of which brutes are innocent. (Of course, one should note the possibility that some animals satisfy the conditions for thick blameworthiness, and it might even be the case that some ordinary-seeming persons fail to satisfy thick blameworthiness while some brutes do.) My proposal is that Smart's framework can accommodate both the thick and thin conceptions just distinguished and should do so.

Smart's position then would seem to be that there will always be a countervailing theoretical reason to eschew the enhancement of responsibility even if there is a pragmatic reason to adopt it. Moreover, the theoretical reason always trumps the pragmatic reason.

This position is subject to doubt on two counts.

First, it is not clear to me that incorporating blame and reproach and their opposites into the practice of responsibility need involve the error of supposing that individuals have libertarian free will. Consider the first-person case. Understanding that I lack free will as the libertarian conceives it, I understand that the only possible moral basis for deep attribution of responsibility is lacking. Whether I do my duty or not, my act will have been caused to be what it is in ways that are beyond my power to control, so any notion of blaming or praising that implies the contrary is tangled in error. But I still

might resolve to heap reproach on myself if my act is a violation of duty, and to regard myself as a less worthy and virtuous person in that event. I respond to my own behavior with reactive attitudes. The rationale for self-blame and self-reproach (and self-praise if my act turns out dutiful) is that knowing that I am so disposed may help to precipitate the causation of a better act than would occur otherwise and that suffering the penalties of adverse self-judgment if I choose wrongly may tend to influence for the better the acts I choose in the future.

Second, even if it were true that adding the judging responses to moral responsibility inevitably commits the individual who engages in judging responses to the false belief that the individual who is judged could have done otherwise than she did and has libertarian free will, it does not follow that adding the judging responses is wrong all things considered. Maybe one cannot regard someone as if they had free will in a way that is emotionally convincing and generates good consequences unless one really does believe, here and now, the person has free will, could really have chosen otherwise than she did. One might find the theoretical cognitive error a less weighty reason than the pragmatic consideration that by shaping responsibility to make room for judging responses we bring about a world in which better acts are done and the world is to that extent a better place. How to weight theoretical cognitive error against practical reasons might seem a difficult, perhaps intractable issue. One consideration that suggests that theoretical error avoidance should not be given much weight in this context is that creation and possession of systematic significant knowledge about the causal order of the world are of greater value than creation and possession of small random facts. In the case at hand, the individual who practices judging-inclusive responsibility can be presumed to

know all about the arguments for the causal thesis and against libertarian free will and to understand how the causal thesis fits smoothly with the scientific world view. Her systematic knowledge need not be defective in any way. Where she trips up is when it comes to particular assertions about particular people. Holding these people responsible for good or bad acts, and blaming or praising them for their acts, she finds herself strongly disposed to believe that these individuals could have done otherwise and were not entirely caused to behave as they did. In reflective moments, when the theoretical arguments she accepts are vividly present in her consciousness, she is disposed to retract these false cognitive claims to which she tends to assent when holding people's feet to the fire and awarding hero badges and when she is otherwise enmeshed in the practice of responsibility. These cognitive lapses might be compared to perceptual illusions such as judging that the straight stick seen in water is really bent. These are cognitive peccadilloes. Any reasonable method of trading off theoretical success and practical moral success will allow that the errors of succumbing to free will beliefs in particular local settings are outweighed by the gains from adopting judgment-inclusive responsibility at least if the practical gains are significant.

I conclude that Smart goes astray in not acknowledging the possibility that the practice of moral responsibility might be enabled to do its job better if judging responses are incorporated into the practice. My own view is that it is not merely logically possible but quite likely that judgment-inclusive responsibility would outperform responsibility shorn of judgment.

With this modification, the Smart account, in a nutshell, holds that responsibility is accountability, and the condition that warrants attributions of accountability is

influenceability. To say that an agent is responsible for an act she has done is to say that she is accountable, that is, a fit object of praise or blame, reward or punishment, depending on its quality. The condition that renders an individual a fit object of praise and blame and so on for what she has done is influenceability. At a first pass, one might regard an agent as influenceable with respect to what she has done if imposition on her of praise or blame and so on for doing it would improve the future by affecting the likelihood that the agent will act in a similar way in the future. But the idea of influenceability needs to be broadened. First, an agent should be regarded as influenceable if either prospective threats and promises of praise and blame and so on if she does a deed would alter the likelihood that she does it or retrospective imposition of praise and blame and so on would alter the likelihood that she would behave similarly in the future. Second, one should regard an agent's doing of something as influenceable if prospective or retrospective imposition of praise and blame and reward and punishment would affect the likelihood either that she herself or others will behave in a similar way in the future. And third, one should acknowledge the possibility that prospective threats of blaming responses and promises of praising responses or retrospective fulfillment of the threats and promises might bring about good consequences by means other than altering people's tendency to repeat the behavior in question—for example, by giving pleasure to those who see virtue rewarded or altering their character in some way for the better. Influenceability then becomes perhaps the main way that it becomes the case that imposing praise and blame and the like would improve the future, but what triggers accountability is that imposing praise and blame and the like on an agent for what she has done would produce good consequences of any sort.

## 5. Utilitarianism and Responsibility

Smart's pragmatic account of moral responsibility has been deemed a utilitarian account and then criticized for its utilitarianism. For example, the Smart view of responsibility, it is said, erects no strict moral bar against the deliberate punishment of those known to be innocent of the offense for which they are being punished. Pointing out this disturbing fact, for some critics, amounts to a *reductio ad absurdum* of this entire approach to moral responsibility.<sup>14</sup>

It is correct to associate the pragmatic approach to moral responsibility and utilitarianism to this extent: Utilitarianism will recommend acts and policies that maximize aggregate utility and only those acts and policies. Regarding the assignment of responsibility for actions and choices, utilitarianism holds that responsibility should be assigned in whatever way is utility-maximizing over the long run. Libertarian free will is strictly a "don't care" for the utilitarian. The pragmatic concerns that support the practice of responsibility according to Smart's account include all the concerns that could possibly matter to a utilitarian who is contemplating what stance on responsibility is best to adopt.

However, it is incorrect that hard compatibilism of the sort Smart espouses implies or presupposes utilitarianism. Hard compatibilism denies that the causal thesis is compatible with moral responsibility except in an attenuated form. This hard compatibilist doctrine of moral responsibility asserts that holding an individual responsible for a choice she makes consists entirely in believing that if good consequences would result from imposing rewards or penalties or praise or blame on an individual for making that choice, that is a good reason for carrying out the imposition.



This doctrine of moral responsibility is compatible with holding that on some occasions moral reasons of a nonconsequentialist kind render it the case that even though the good consequences that would flow from imposition of rewards and penalties constitute a good reason for doing so, all things considered the imposition should not be done.

This doctrine of moral responsibility is also consistent with holding that fundamental level moral principles assign moral rights to individuals and that good consequences, including the good consequences that warrant attributions of responsibility, consist in whole or in part in fulfillment of these moral rights. That is to say, it is no part of hard compatibilism that the desirable states of affairs, production of which warrants ascriptions of responsibility, are identified with utility or human well-being or anything of the sort.

Hard compatibilism in fact carries no commitment to any form of consequentialism, so the defects of consequentialist morality, if such there be, are not to be laid at the door of this doctrine on the nature of responsibility.

All that hard compatibilism requires is that morality contains a consequentialist element or component, such that the fact that holding someone responsible would produce best consequences all in all is at least a, not necessarily a decisive, reason for judging that one ought to do the holding. But this is hardly controversial. Hard compatibilism is controversial not for what it includes as relevant to ascriptions of responsibility but for what it excludes.<sup>15</sup>

Some confusion on this point is evident in the literature on free will and responsibility. Against hard compatibilism as espoused by Smart it is objected that this

approach implies that one can be morally responsible for an act one has not even performed and justly punished when one must be deemed innocent on any sensible criteria. If it would produce good consequences, say by deterring crime, to frame Smith, who is innocent, and bring it about that he is punished for a crime he has not committed, then the Smart doctrine of responsibility must endorse this result. So it is claimed.

But if we turn back to Smart's essay, we find that Smart does not commit himself to this result. No doubt the position that the judicial punishment of the innocent is morally right and ought to be done just in case it would produce better consequences than any available alternative is one that Smart, a committed utilitarian, endorses. But the hard compatibilist position on moral responsibility that Smart develops in the essay we are reviewing does not in and of itself yield this result. Illustrating his approach to responsibility, Smart invents a schoolteacher who is faced with a student who performs poorly. The poor performance might be due to laziness or lack of intelligence. If the latter, Smart assumes, then no good will come of imposing penalties for poor performance. If the former, then there might well be a point to penalizing the lazy student. Performance might improve. The student's performance will be brought about by prior causes in any event, but the teacher's threats can be part of the causal environment that determines the student's choice.

Notice that Smart does not consider the wider possibility that punishing the unintelligent boy might produce good consequences by bringing about improvement in the performance of other students, who may believe their performance will be more closely monitored if punishment of their mate occurs. The more rigid regime of rewards and penalties that is responsive only to the level of students' performance and does not

discriminate between (1) performance caused by factors that threats of punishment and promises of award can influence and (2) performance not influenceable in these ways may more reliably spur efforts at performance because the student then does not have the option of intentionally bringing it about that he appears to satisfy one of the excusing conditions. But Smart does not raise this possibility. Nor does he consider the further possibility that falsely accusing Jones (who was not enrolled in school at the time) of poor performance on the test and imposition of exemplary strict punishment might effectively terrorize the students into best performances or in some other way induce better consequences than declining to punish Jones unjustly.

The strategy Smart follows in developing an account of moral responsibility might be described as conservative. He wonders what remnant of the idea of responsibility is sustainable given the likely fact that human choices are events and all events are physically caused either deterministically or indeterministically but at any rate in a way that leaves no room for free will and agency as the libertarian wishes to conceive them. The remnant he discovers is the idea that one is responsible for an act if one did it and doings of this sort are influenceable by blaming or punishing. One cannot squeeze hard on this admittedly thin notion of responsibility to somehow induce it to imply that one can be responsible for a crime one did not commit, because one's doing it is by definition required for responsibility.

Of course if one proceeds to ask Smart, who happens to be a utilitarian, what one ought to do if one were faced with a scenario in which by acting exactly as though one were blaming or punishing Smith for a misdeed that one knows he did not actually do one could bring about a greater good than one could achieve by any alternative available

choice one has, his reply is obvious. But the answer to this sort of question has no more tendency to somehow force expansion of the thin notion of moral responsibility than would the answer to the question, what if one could maximize human happiness by telling a lie, somehow provide grounds for altering our conception of a lie.

### **6. Voluntariness and Responsibility**

Further support for Smart's position on responsibility emerges from reflection on cases to which Robert Adams draws attention.<sup>16</sup> Anger without just cause and smoldering resentment that outlasts proper occasions for forgiveness can be morally blameworthy. Such attitudes are morally unjustified, and they often cause significant harm. But in many cases they are not the product of voluntary choice, and in fact are beyond voluntary control. Anyone who insists that we can only be morally responsible for what lies within our power to control must then deny that an agent can be morally responsible for having morally bad attitudes when having them does not lie within her power to control. One might try to hold onto the claim of moral responsibility for having morally bad attitudes by tracing back the inculcation of the attitude to a choice of the agent that if made rightly would have blocked its inculcation. When this occurs, the agent could be responsible for having the bad attitude even though she cannot control its expression now because in the past a choice she should have made would have prevented her from developing the attitude. But in many cases looking back to the past in this way will not reveal any prior act of choice that can serve as the voluntary-control basis for responsibility.

Adams notes that involuntary sins of the heart often cause harm through their involuntary behavioral expression. He clearly rejects the reduction of responsibility to

influenceability and holds that one can be responsible and blameworthy for what one is as well as for what one does even in some kinds of case in which what one is does not lie within one's voluntary control. The ultimate basis of responsibility on his view is not that treating people as though they were responsible is productive of good consequences.

However, my sense is that one's willingness to accede to Adams's judgment that the agent is responsible for having morally bad attitudes in the sorts of cases he describes is driven entirely by the consideration that taking responsibility for one's attitudes and holding other people responsible for theirs might improve the world. It does sound callous of me to say of my own self-righteous anger that this is the involuntary expression of a disposition that was instilled in me by forces beyond my power to control and triggered by circumstances that were not at all of my making. The impression of callousness made by this remark is not dissipated by the assurance that what I am saying is one hundred per cent true. Where is the toehold for an attribution of responsibility here?

My surmise is that what triggers unease about such a disavowal of responsibility is the hunch that it reinforces the causal factors that have produced and sustained the morally repulsive attitude. By disavowing responsibility for one's own unjust anger one fosters it, whereas disavowing any responsibility for a stranger's unjust anger has no comparable tendency to bring about evil consequences. Hence the former disavowal can be morally wrong and the latter not, given their different consequences.

This way of viewing the examples gains support if we instead imagine a case in which the agent's assumption or nonassumption of responsibility for her bad attitudes is causally inert. Suppose a raft of soon-to-be-discovered psychological research should

convince any rational person that his anger (for example) is entrenched in such a way that mental fretting about it would have absolutely no power to reduce its incidence or bad effects. Then one should not fret about it. Nor should one rehearse acts of contrition or repentant avowals of responsibility coupled with sincere resolve to try to change.

In another sort of possible case, beating one's breast and acknowledging responsibility for one's involuntary sins might be counterproductive in the sense that avowal of the bad attitude entrenches it. Here it would be better to disavow responsibility.

In many cases one may not know whether the assumption of responsibility for one's attitudes will be productive, counterproductive, or inert. Here what should be said is that if what a person ought to do depends on facts she does not know, then she will not know what she ought to do. This scenario is not an embarrassment for the influenceability account of responsibility.

Holding oneself morally responsible is subjecting oneself to self-administered sanctions. According to the influenceability account, this subjection is undesirable if it issues in no good consequences. Holding oneself responsible involves using one's moral responses to one's own conduct as a tool to produce results—hence the external aspect of which Wallace complains. It might be that disposing oneself immediately to feel guilt in response to one's own conduct that fails to meet an accepted standard without thinking about the further consequences would produce better consequences than attempting to dole out blame and reproach to oneself as a result of instrumental calculation at the time of the self-imposition. If so, the influenceability account may endorse the noncalculating disposition. We should distinguish what holding oneself responsible amounts to and

what one should have in mind when reproaching oneself in the course of holding oneself responsible.

It is interesting to note that even though Adams explicitly rejects the idea that moral responsibility reduces to influenceability, he is careful to avoid endorsing undue reproach when it is directed at uninfluenceable bad attitudes. If Smith is trying unsuccessfully to alter his entrenched bad attitudes, the virtue of mercy counsels observers not to heap reproach on him when that “serves no good purpose.” Of course Adams is free to hold and does hold that reproach can be deserved even when no one ought to express it. Adams’s account of responsibility for bad attitudes not under the agent’s voluntary control, once its details are in place, rarely requires him to assert what the influenceability theorist will want to deny concerning the occasions when holding someone responsible is justified. If one focuses specifically on cases in which a person evinces a bad attitude ineluctably, then if it is also the case that no imputation of responsibility to this person in any form by anybody will do any good, the impulse to endorse any imputation of responsibility in this sort of case should be firmly resisted.

### **7. Influenceability versus Responsibility**

Objections to the influenceability account may seem to protrude from Scanlon’s correct observation that there is not “a good fit between the idea of influenceability and the idea of responsibility which we now employ.”<sup>17</sup> The problem that Scanlon puts his finger on is that many factors other than the nature of the agent’s act play a role in determining whether it is useful to praise or blame it.

For example, imagine that Mafia thug terrorizes a small village. He commits many heinous crimes. But as it happens any attempt to punish or reproach him will be

unsuccessful, will only make him irritable, and hence will lead him to act more brutally. Even self-reproach would have no effect other than to make him more prone than he otherwise would be to angry, immoral outbursts. The influenceability theory then must say that he is not morally responsible for his misdeeds, which seems odd, for no standard excuses exempt him from blame. For all that has been said so far, he might have libertarian free will.<sup>18</sup>

The invocation of libertarian free will changes the picture by eliminating the circumstances that give point to the influenceability theory. So we should set this possibility to the side.

Still, the reasons that can render it sensible to hold someone responsible for a misdeed may be extraneous so far as its quality is concerned. In the circumstances described, holding the Mafia guy responsible for his crimes would be mistaken, whereas holding responsible and punishing a mentally retarded and mentally ill person who entirely lacks rational agency capacity might yield good consequences and so be justified according to the influenceability account.

Holding an individual responsible in the sense of liable to praise or blame for an action or omission might produce good consequences by way of its impact on the individual herself, on other people who might be induced to alter their behavior or attitudes, or on the agent who engages in responsibility ascription. We should probably toss into the hopper the satisfaction gained by those who experience vicarious reward if the doer of the good deed is rewarded (the satisfaction of those who gloat at punishment is probably a disvalue). All of these consequences might accrue either in the short or the long run.



Influenceability has to be regarded as a substitute for the ordinary idea of responsibility, according to which only the quality of will of the agent manifested in a particular choice or omission matters for the assignment of responsibility for that choice or omission. A precondition of finding influenceability acceptable is having good grounds for finding the ordinary notion of responsibility unsustainable. So the fact that influenceability does not mesh perfectly with the ordinary notion of responsibility is not per se an objection to it.

Regarding the Mafia thug who is impervious to influence, one can say that his deeds are morally heinous. If he cannot be influenced for the better by the assignment of responsibility (and no other good consequences are in the offing), no assignment should be made. We should note a limit case here: it might be that victims and observers of the thuggery can do good by blaming the perpetrator in their hearts even if no external expression of such blame is warranted on consequential grounds.

### **8. Responsibility and Desert and Theories of Justice**

Issues regarding free will and moral responsibility are intertwined at the center of recent discussions of distributive justice.<sup>19</sup> Does justice fundamentally require that people get what they deserve?

Consider a health care agency that strives to improve the health of people who cannot afford to pay for medical care or medical insurance out of their own pockets. Suppose for simplicity that the goal of the agency is thought to be to save as many as possible of the lives of those whom disease and disability and accident have placed at risk of premature death. Distributive justice as some conceive it suggests that the agency's goal should be reformulated. Some people find themselves at risk of suffering premature

death because their genetic susceptibilities have played themselves out in ways that are entirely beyond their power to control. Other people face similar risks that have materialized entirely as a result of their voluntary choices to engage in risky activities such as smoking or driving after drinking or engaging in hazardous sports. The responsibility that choice entails need not involve any imputation of fault. Someone might have made a choice that is, all things considered, perfectly reasonable, but provided she made the choice freely from a wide range of alternatives including ones that would have offered satisfactory valuable outcomes, it is reasonable to hold her responsible for her choice and for the outcomes that accrue as a result to others and to herself. Responsibility-catering doctrines of distributive justice suggest that the moral value of obtaining a benefit of a given size for a badly off person increases, the less she is reasonably held to be responsible in virtue of her prior choices for her present plight (if it is bad). Responsibility-catering doctrines also hold that if a person's situation is good, so that her lifetime well-being is high, bringing it about that she suffers a well-being loss in order to benefit other people is intrinsically morally less desirable, the more she is responsible for her present good fortune.<sup>20</sup> These doctrines could just as well be formulated in the language of desert: Other things being equal, it is better to bring it about that good fortune goes to the deserving.

The Smart theory of responsibility rejects the idea that individuals are ever truly virtuous or vicious. Nor can they be responsible or irresponsible for what they do in any deep sense. (In a shallow sense, people can be responsible for what they do in so far as holding them responsible would be productive of good consequences.) This being so, it is not intrinsically morally valuable that the virtuous and responsible enjoy good fortune

to a greater extent than the vicious and irresponsible. If we must allocate a good or evil to someone, the virtue status and responsibility ranking of the person do not affect the intrinsic moral desirability of the state of affairs in which she gets the good or evil. If there were true virtue or truly responsible choice or true desert, it might be morally a good idea to reward it, but since these concepts have no application, in effect the virtue and responsibility and desert rankings of any person are always the same as anyone else's, so moral principles that specify that the good and bad fortune that people get should vary with the moral quality of their choices have no application. Punishing the Hitlers of the world might prevent holocausts and other evils and rewarding the Mother Therasas of the world might feed the poor of Calcutta and deliver other good effects. If so, there is reason to do the punishing and rewarding. But that is all there is to responsibility. It is simply a tool of social policy and self-culture.

### **9. Soft Compatibilist Alternatives to Hard Compatibilism**

In “Free-will, Praise, and Blame” Smart does not mention the possibility that there might be alternative conceptual perches between hard compatibilism and libertarianism. In recent years, a multiplicity of soft compatibilist perches have been thoroughly explored. Although Smart suggests no arguments that undercut soft compatibilism, the coherence and plausibility of the doctrine are still disputed.

Whether there is a viable basis for attribution of responsibility other than libertarian free will is an issue lurking in the background of recent controversies concerning social justice and personal responsibility.

For example, consider discussions between Dworkin and his critics on Dworkin's treatment of responsibility.<sup>21</sup> Dworkin approaches responsibility in the course of

elaborating an account of equality of resources, his preferred ideal of justice, the details of which do not concern us. Dworkin proposes that individuals should be responsible for their option luck, in the sense that the effects of an individual's option luck that fall on herself should not trigger a claim for further compensation in the name of justice. Other people are under no obligation of justice to make good losses for an individual that stem from her option luck. In contrast, losses that arise from brute luck should be undone, and gains that fall on an individual as a result of brute luck might be redistributed to others without counting as an injury to her. Dworkin stipulates that "[o]ption luck is a matter of how deliberate and calculated gambles turn out--whether someone gains or loses through accepting an isolated risk he or she should have anticipated and might have declined. Brute luck is a matter of how risks fall out that are not in that sense deliberate gambles."<sup>22</sup> Here Dworkin is assuming that a fair framework of interaction is in place.

Dworkin holds that individuals are responsible for their choices but not for chance happenings—brute luck fortune and misfortune that fall on a person in a way that is not mediated by her choices. He adds that individuals are responsible also for what underlies their choices, their aims and ambitions that form their conception of what is worth striving for. These are aims and ambitions that the individual endorses, is glad to have. These are contrasted with mere cravings that an individual experiences but that do not form part of her conception of her good and are just distractions or obstacles to its attainment. (An individual might acquire a craving as a result of option luck, in which case she would bear responsibility for having it, but genetics and socialization might just impose a craving on her.)

The critic objects that the line between ambitions and cravings does not coincide with the divide between choice and chance. It might be a matter of sheer brute luck (for which an individual cannot reasonably be held responsible) that she comes to have a particular aim or ambition. Hence the individual should not be held responsible for having this ambition, and this judgment of nonresponsibility should qualify or perhaps negate the claim that the individual should be deemed responsible for any choice she makes to achieve this ambition.

Contrast the willing addict, the righteous dope fiend who endorses drug taking, with the unwilling addict, who does not value drug taking and wishes she could be rid of the desire that induces her to engage in this activity. The critic holds that the process by which the individual came to endorse drug taking may have been beyond her power to control. If so, she should not be held responsible for the endorsement.

Dworkin demurs. He urges that there is no room for ever holding people responsible for their choices and aims while declining to hold anyone responsible for some choices and aims on the ground that the latter do not lie within the agent's power to control. The criterion for responsibility the critic must be proposing is too sweeping. It eliminates all responsibility, or at least must do so in a causally ordered world in which all that happens is ultimately fixed by prior causes beyond any agent's power to control.

Now one might be inclined to protest that this response to the critic is far too swift. What is supposed to bring it about that there is no conceptual room for an account of responsibility for aims and ambitions more demanding than the glad-to-have criterion that Dworkin proposes?

Consider two persons who embrace the same low-grade aims and ambitions and in their pursuit make comparably defective choices about how to live their lives and end up with the same poor quality of life. The two persons might be differentially responsible for their endorsement of defective aims. Suppose that avoidance of this endorsement requires a sophisticated appraisal of complex arguments, and that this appraisal would be very difficult for Smith, who is unintelligent through no fault of his own, and easy for Jones, who is intelligent by good fortune. Smith conscientiously engages in critical reflection but cannot get the right answer, while Jones lazily foregoes reflection. Or suppose that avoiding embracing defective values would be very painful for Smith, due to peer pressure, but not unpleasant for Jones, who faces no painful hurdle to sound value selection. At the limit, making a reasonable value choice might be so difficult and painful as to be impossible for Smith but readily possible for Jones. Short of that limit, making a reasonable value selection might be very difficult and costly for Smith but not for Jones. Hence we might judge Smith to be not responsible at all or (short of the limit) less responsible than Jones for the bad values she embraces.

Dworkin's defense of the glad-to-have account of responsibility for ends might seem to fail on the ground that it dismisses plausible alternative accounts without argument. But I think he has a point.

We expect someone to resist an unworthy motive. But resistance must also have a motivational source, which will not be voluntarily chosen by the individual. Or even if there is an element of choice in the formation of this motive, further back in the causal chain, unchosen proclivity will inevitably emerge. If one person repents of his sin while another does not, and the first person chooses to repent, there will be a pang of guilt or

some other trigger to choice of repentance, and the pang or whatever will not itself be voluntarily chosen. Nor will it be within the voluntary control of the individual. The same will be true for any pair of individuals who embrace a defective value, of whom one is claimed to be alone responsible for the embrace. If one engages in reflective scrutiny while the other does not, some desire prompts the one who engages to engage, and even if this desire is chosen, some antecedent desire will simply occur unchosen.

Dworkin thus maintains that if we insist that one can reasonably be held responsible only for what lies within one's power voluntarily to control, we are setting the bar too high. Judged by that high standard, it will turn out that no one is responsible for anything. But if we do not accept that excessively demanding standard, we have no basis for rejecting the glad-to-have criterion for responsibility for ends, which anyway fits our commonsense practice of assignments of responsibility.

As stated, Dworkin's argument does not rule out a libertarian standard of responsibility. But if we eliminate libertarian accounts on the ground that they probably impose conditions for responsibility that are never satisfied in the world we inhabit, as we see if we take our bearings from the current trajectory of science, then Dworkin's riposte to the critic is powerful.

The riposte denies that there is conceptual space for a more robust account of responsibility for ends than the glad-to-have account. This riposte does not offer to assuage the critic's grounds for doubting the adequacy of the glad-to-have account. The point is rather that the critic's objections would if accepted lead inevitably to the repudiation of all soft compatibilist accounts of responsibility, according to which individuals can be truly praiseworthy and blameworthy, truly responsible and

irresponsible, such that it is intrinsically morally desirable that some individuals should be rewarded, some punished, some made to enjoy higher and some a lower quality of life. This line of thought leaves entirely open the possibility that soft compatibilism indeed deserves to be abandoned.

The disagreement within the theory of justice between Dworkin and his critics echoes the disagreement within recent free will discussions between advocates of mesh and historical accounts of responsibility.<sup>23</sup>

The mesh advocate holds that whether a person should be held responsible for a choice that she makes depends entirely on structural relations among that factors that led to her choice and not at all on the history of how that web or mesh of structural factors came to be. The historical advocate insists that whether an individual is responsible for a choice that she makes depends in part on the history or the process by which the mesh singled out by the mesh theorist came to be (if indeed any of the conditions of responsibility refers to such a mesh), and in any event whether an individual is responsible for a choice depends on the character of the process by which the choice originated.

The glad-to-have account of responsibility for ends is an example of a mesh theory. Whether an individual is responsible for an end or ambition she has depends on whether she wholeheartedly embraces it or endorses it. The mesh in this instance is a fit between the end held by the individual and her attitude of endorsement of it. How the endorsement came about does not bear on the issue of responsibility according to this account.



The historical theorist objects to any theory of responsibility of this type. The objection is that whatever is proposed to constitute the appropriate mesh of factors that gives rise to responsibility, it might come about by a process that is unequivocally responsibility-undermining. The agent might undergo some psychological process such as brainwashing and as a result come to exemplify the mesh, but in virtue of the origin of the agent's state, it should not be thought to render the agent responsible for the choice or end in question. The point is made sharply if one imagines intervention in the agent's process of attitude and belief formation by aliens from outer space who control the agent's mental state by beaming rays at her. An agent is rightly deemed responsible for her choices and for the ends she seeks to fulfill only if the process by which she embraces her ends and makes her choices has the appropriate character that is not subversive of responsibility. This being so, says the historical theorist, no mesh account of responsibility can be the whole story specifying sufficient conditions for responsibility.

The rejoinder offered by the mesh theorist parallels Dworkin's response to his critics. To insist that no condition of an agent can be sufficient to establish the agent's responsibility for her ends and choices unless the history of the origination of this condition is appropriate is to set the standard of responsibility too high and to insist on requirements that cannot be satisfied at least in a world where everything that occurs is an event brought about by prior causes. In a world of cause and effect, any account of how an agent's embrace of her ends and making of her choices originates will eventually trace back to prior sufficient conditions (or sufficient modulo brute statistical regularity) that existed before the individual was born and for which she could not possibly bear responsibility.

The historical theorist will hold that just as not all causes are created equal, not all histories are created equal either. According to an historical compatibilist account of responsibility, not all causes of choices undermine the agent's responsibility for them, just some special sorts. And in a similar way not all histories undermine the agent's responsibility, just some special sorts.

The dialectic of argument between the historical theorist and the mesh theorist indicates a line of thought that supports the abandonment of any soft compatibilist account of moral responsibility. One simply puts together two claims made by the historical theorist and two by the mesh theorist as follows:

1. No mesh theory of responsibility for an agent's selection of ends and making of choices is correct, because any such theory sets no limits on how it comes about that an agent satisfies the conditions necessary and sufficient for responsibility as stipulated by the theory.

2. For any mesh theory, some ways that individuals might come to satisfy the conditions for responsibility that the theory stipulates are themselves responsibility-undermining and hence incompatible with responsibility.

But

3. No historical account of an agent's responsibility for her selection of ends and making of choices is compatible with holding that any agent is ever responsible for her selection of ends and making of choices.

This is so because

4. The history of how it comes to be that any agent selects her ends and makes her choices eventually traces back to conditions existing prior to her birth, for which she

could not possibly be held responsible, and that are causally sufficient (up to indeterminacy due to chance) for bringing it about that the agent selects the values and makes the choices that she does.

I do not offer this argument as a substantive contribution to the debate concerning the viability of soft compatibilism in its best elaboration. The mesh theorist will have reasons for denying 2 above and hence 1, while the historical theorist will have reasons for rejecting 3 and 4. I simply wish to indicate that Smart's inference from the rejection of libertarian free will to the position that only hard compatibilism is viable is not clearly mistaken despite his failure to consider the possible varieties of soft compatibilism. For at the end of the day it might be the case that none of these varieties is viable.

#### **10. Attributions of Responsibility and Moral Reasoning Capacity**

A prominent strand of recent soft compatibilist thinking emphasizes possession of moral reasoning ability by a person as the basis for attributing responsibility to that person. In separate writings Gary Watson and Jay Wallace have elaborated this idea persuasively.<sup>24</sup> In Wallace's formulation, "it is fair to hold people morally responsible if they possess the rational power to grasp and apply moral reasons, and to control their behavior by the light of these reasons."<sup>25</sup> A morally responsible agent on this view is able to respond to moral criticism of her behavior by considering and assessing the reasons for and against it and eventually arriving at a considered moral judgment as to its appropriateness. Moreover, the responsible agent is capable sometimes of adjusting her behavior to bring it into conformity with the verdicts of moral deliberation.

The hard compatibilist can agree that many humans have the capacity, to varying degrees, to engage in critical moral reflection. She can agree that our ordinary practice of

moral responsibility involves holding people to norms by engaging them in moral deliberation concerning their own and other people's conduct. She casts her net more widely in maintaining as well that holding responsible individuals such as the retarded and insane who lack moral deliberation capacity by praising and blaming and rewarding and punishing them for their deeds might (conceivably) also make sense if holding them responsible in this way increases desirable conformity to norms. (Whether such treatment of moral incompetents would be morally justifiable all things considered is a question that is not settled by the idea of moral responsibility itself, but by whatever moral principles bear on this issue. The hard compatibilist takes no stand on this issue beyond denying that the notion of moral responsibility itself forbids treating moral incompetents as morally responsible.)

The hard compatibilist will also cast the net of responsibility somewhat more narrowly than Wallace, in that she will not place much weight on the distinction between having the capacity to engage in moral reasoning and be affected by it and having the disposition so to engage and be affected. If a psychological test reveals that an individual has the ability to engage in moral reasoning, but he is irrevocably disposed not to engage, so that the panoply of moral reproach, reward, and punishment applied to him does no good, whether by altering his behavior or anyone else's, then it is unfair to impose costs on him in the name of holding him morally accountable to no good purpose.

In another respect the hard compatibilist can find room for agreement with Wallace's soft compatibilism. We engage in moral reasoning with others in order to clear our heads and make progress toward clarifying what is really morally right and wrong. To engage with others in this enterprise might be called treating them as morally

responsible agents. With all of this the hard compatibilist has no quarrel. She would merely insist that having and exercising the capacity to engage in moral reflection does not render anyone truly morally blameworthy and praiseworthy for her conduct in a deterministic world. In such a world true blameworthiness and praiseworthiness are not to be had, so attributions of moral responsibility that involve the judgment that agents can be truly blameworthy and praiseworthy should be scuttled. In a deterministic world what renders agents praiseworthy and blameworthy in a sense, fit objects of reactive attitudes, is that their being subject to praise and blame and the like improves the future.

According to Wallace a further condition that must obtain if holding someone morally responsible is to be fair is that besides having the capacity for moral reasoning, the individual must be able to control her behavior in the light of these reasons. For the soft compatibilist, the control in question is conceived to be consistent with determinism.

The hard compatibilist agrees and disagrees. First, consider an agent for whom moral reasoning is an idle wheel. He can alter his behavior only in response to prudential reasons. This fact in itself does not bring it about that holding him morally responsible for his conduct makes no sense or must be unfair. After all, the individual is influenceable. Promises and threats of praise and blame, reward and punishment affect his behavior. The notion of holding the individual responsible for his conduct is then at home, according to the hard compatibilist. If one objects that this morally incompetent agent cannot be truly guilty of crimes he commits, one is implicitly dividing those who violate moral norms into those who are truly blameworthy and those who are not, but the hard compatibilist regards this contrast as spurious.

Second, consider an individual who is able to control his behavior in the light of moral reasons. In a deterministic world, this means that the perception of good reasons often causes him to act as the good reasons dictate. Now take a case in which this ordinary rational agent violates a moral norm. The causal mechanism that normally issues in conformity to moral norms misfires in this case. The hard compatibilist will say that such an agent is likely to be influenceable by moral sanctions, so holding him responsible makes sense in this case. But ultimately it is misleading to speak of the individual as having an ability to control his behavior which he exercises in a wrongful way on this occasion. This way of speaking conjures up associations of contracausal freedom which the hard and soft compatibilist both agree are out of place on the assumption the world is deterministic or causally governed. The hard compatibilist should deny that there is some deep difference between the ordinary rational agent and the moral incompetent that brings it about that attributions of moral responsibility are fair only as applied to the former.

Of course, given determinism, to conceive of individuals as agents who choose among alternative courses of action and act freely is already to be on a slippery slope, at the bottom of which lie paradox and contradiction. Determinism unsettles familiar and deep-seated notions. But this fate is common to hard and soft compatibilists. They just differ about what is all things considered the best strategy for conceptual and normative damage control.

## **11. Conclusion**

This essay has criticized Smart's premature dismissal of libertarianism. But taken on its own, the account of moral responsibility as influenceability, I claim, is less

counterintuitive than is often supposed. Insofar as this doctrine is counterintuitive, what is unsettling is required by the most sensible and plausible normative response to the conjecture that human choices are caused events.

---

<sup>1</sup> . J. J. C. Smart, “Free-will, Praise, and Blame,” *Mind*, vol. 70 (1961), pp. 291-306.

See also Moritz Schlick, *Problems of Ethics*, tr. by David Rynin (New York: Prentice-Hall, 1939), chapter 7, pp. 143-158.

<sup>2</sup> . The thesis that threatens free will is that all human choices are events and events are caused by prior events according to physical laws. The causation here might be deterministic or indeterministic, the latter if the fundamental physical laws are statistical. I sometimes refer in the text to a deterministic world to invoke the idea that human choices along with all other events are caused according to physical laws (whether the laws are deterministic or not).

<sup>3</sup> . R. Jay Wallace, *Responsibility and the Moral Sentiments* (Cambridge: Harvard University Press, 1994), p. 57.

<sup>4</sup> . Saul Smilansky, *Free Will and Illusion* (Oxford: Clarendon Press, 2000), p. 33.

<sup>5</sup> . T. M. Scanlon, “The Significance of Choice,” in Sterling M. McMurrin, ed., *The Tanner Lectures on Human Values*, vol. 8 (Salt Lake City: University of Utah Press, 1988), p. 160.

<sup>6</sup> . Smilansky’s position is complex, and is not accurately describable as soft compatibilist.

<sup>7</sup> . The yellow stain discussion occurs in J. J. C. Smart, “An Outline of a System of Utilitarian Ethics,” in Smart and Bernard Williams, *Utilitarianism For and Against* (Cambridge: Cambridge University Press, 1973), p. 52.

---

<sup>8</sup> . Smart, p. 306.

<sup>9</sup> . Shelly Kagan explores views of this sort in "Equality and Desert," in Louis P. Pojman and Owen McLeod, eds., *What Do We Deserve?* (Oxford: Oxford University Press, pp. 298-314.

<sup>10</sup> . David Miller, *Principles of Social Justice* (Cambridge and London: Harvard University Press, 1999), chapters 7-9.

<sup>11</sup> . Smart, "Free-will, Praise, and Blame," p. 300.

<sup>12</sup> . The classic account of the role of reactive attitudes in the practice of moral responsibility is Peter Strawson, "Freedom and Resentment," reprinted in John Martin Fischer and Mark Ravizza, eds., *Perspectives on Moral Responsibility* (Ithaca and London: Cornell University Press, 1993), pp. 45-66.

<sup>13</sup> . Discussing the Schlickian account of responsibility, Jonathan Bennett writes that it "cannot do justice to the real nature of our praise- and blame-related responses. When we express indignation for someone's cruelty, or admiration for his unselfishness, we usually are not engaged in any sort of therapy: blame-related responses all involve something like hostility toward the subject." See Bennett, "Accountability," in *Philosophical Subjects*, ed. Zak van Straaten (Oxford: Clarendon Press, 1980), pp. 14-47; see p. 20. I do not see why an account that holds that people should be held accountable when the consequences of doing so are desirable cannot help itself to the idea that accountability should include reactive attitudes when their inclusion makes the consequences better. For this to work, the agent at the time of praising and blaming probably cannot have in mind the thought that she is behaving strategically to induce good consequences. But the conditions that warrant accountability need not be in the mind of someone engaged in an accountability



---

practice. After all, Bennett following Strawson holds that praise and blame and the like are natural human reactions; the question is just whether they should be inhibited or encouraged under some circumstances.

<sup>14</sup> . See the discussion in Smilansky, *Free Will and Illusion*, at pp. 27-33.

<sup>15</sup> . Discussing Hilary Bok's views, Derk Pereboom maintains that conceptions of responsibility such as responsibility-as-influenceability do not engage the debate about free will and responsibility. He writes that "the issue in the debate about freedom and moral responsibility is not whether determinism threatens the legitimacy of holding oneself and others morally accountable, if, for example, this legitimacy just consists in the fact that it would be effective for moral improvement. See Hilary Bok, *Freedom and Responsibility* (Princeton: Princeton University Press, 1998); also Derk Pereboom, *Living without Free Will* (Cambridge: Cambridge University Press, 2001), p. xxi.

<sup>16</sup> . Robert Adams, "Involuntary Sins," *Philosophical Review*, vol. 84 (1985), pp. 3-31.

<sup>17</sup> . Scanlon, "The Significance of Choice," p. 160.

<sup>18</sup> . Fred Feldman suggests an example that sharpens the point. Suppose Smith maliciously presses a button that unleashes an explosion that will destroy all life on earth in two seconds. Blaming him even in thought could produce no good consequences, so according to the account of responsibility I favor, Smith is not responsible for his misdeed. Given that libertarianism is ruled out, I would hold that this implication of the account is acceptable. It should be noted that there are views in the neighborhood of Smart's position that could deliver the response that Smith is responsible in this example. For example, one might hold that one is responsible if one is influenceable and one is influenceable if under normal circumstances the imposition of hard treatment or blaming

---

responses or the like would affect one's tendency in the future to repeat the behavior in question. If Smith's behavior can be altered by sanctions under normal circumstances, then he is influenceable, and responsible, even if in this case circumstances are abnormal. Shelly Kagan pointed out this possible line of argument.

<sup>19</sup> . See Robert Nozick's criticism of John Rawls's theory of justice on this score in *Anarchy, State, and Utopia* (New York: Basic Books, 1974), chapter 7, section 2.

<sup>20</sup> . I discuss responsibility-catering approaches to distributive justice in "Luck Egalitarianism and Prioritarianism," *Ethics*, vol. 110 (2000), pp. 339-349.

<sup>21</sup> . Ronald Dworkin, *The Sovereign Virtue: The Theory and Practice of Equality* (Cambridge and London: Harvard University Press, 2000), chapter 7. The critic discussed by Dworkin is G. A. Cohen, "On the Currency of Egalitarian Justice," *Ethics*, vol. 99 (1989), pp. 906-942.

<sup>22</sup> . Dworkin, *Sovereign Virtue*, p.

<sup>23</sup> . The discussion below of historical and mesh accounts of moral responsibility is indebted to John Martin Fischer and Mark Ravizza, *Responsibility and Control: A Theory of Moral Responsibility* (Cambridge: Cambridge University Press, 1998). The terminology of "mesh" versus "historical" accounts is theirs.

<sup>24</sup> . For Watson's views, see his "Free Agency," *Journal of Philosophy*, vol. 72 (1975), pp. 205-220; Watson, "Soft Libertarianism and Hard Compatibilism," *Journal of Ethics* 3 (1999), pp. 351-365; also Watson, "Reasons and Responsibility," *Ethics*, vol. 111 (2001), pp. 374-394; see esp. p. 393.

<sup>25</sup> . R. Jay Wallace, *Responsibility and the Moral Sentiments* (Cambridge and London: Harvard University Press, 1994), p. 7.