

On Responsibility in Science and Law

John Staddon¹

Duke University and the University of Western Australia

Respon'sible, liable to be called to account or render satisfaction: answerable: capable of discharging duty: able to pay.² The old Chambers's dictionary gives a behavioristic view of responsibility: in terms of action, not thought or belief. "Lust in the heart" is not equated to lust *in flagrante*. It is this view I shall explore in this paper, rather than the more subjective notion of *moral* responsibility, as in "I feel moral responsibility (i.e., *guilt*) for not doing anything to save the Tutsis [Hutus, ethnic Albanians, etc.]." My presumption is that responsibility implies capability: you cannot be held responsible for something over which you have no control (Hutus, ethnic Albanians, etc.). There is obviously uncertainty in some cases – where control is less than total, where the degree of potential control depends on our own efforts, where we can't know if we have control until we try to exert it, etc. The relation between responsibility and degree of control is a separate topic I shall not treat here.

Responsibility in the behavioristic sense is also *social*, we are responsible *to* some person or group,³ and therefore goes beyond the individualistic focus of most psychology and psychiatry. I equate "behavioristic" with "scientific" because science is generally agreed to deal only with "third-party-confirmable" concepts, "public knowledge" in Ziman's phrase⁴. I shall contend that responsibility in law traditionally has been, and should be, behavioristic, concerned with intersubjectively verifiable actions. I argue that responsibility and determinism are not antithetical but mutually supportive ideas; that factors affecting responsibility, such as drugs and mental and physiological conditions, may be the occasion for increased or decreased penalties; and that the decision in such cases is not scientific but moral. I conclude, *contra* some modern authorities, that there is no opposition between science and law.

Responsibility and Determinism

Generations of introductory psychology students have learned that human behavior is governed by causal laws, that it is *determined*. Their teachers often go on to assert that responsibility and determinism are mutually exclusive, that if a man's⁵ behavior is entirely caused by his personal history (especially if that history is tragic or aberrant) he cannot be held fully responsible for his actions. Several observers have noted that this idea of diminished responsibility is usually not extended to positive accomplishments. Great achievements are rarely dismissed as the inevitable byproduct of a felicitous upbringing, even though failures and misdeeds are frequently excused by a bad one. The opposition of determinism to responsibility, promoted persuasively by psychologists of all stripes, from radical behaviorist

¹ Send reprint requests to the author, Department of Psychology: Experimental, Duke University, Durham, NC 27708. E-mail: staddon@psych.duke.edu.

² *Chambers's Twentieth Century Dictionary*, 1901/1950. There is very little new that can be said on the topic of this paper – every idea has its antecedents and a paper that attempted to cite them all would be impossible to write, to read – and to print. I have done my best to reference key items, taking comfort from Alfred North Whitehead's aphorism "Everything of importance has already been said by someone who did not discover it."

³ Some talk of "responsibility to oneself", but this is metaphoric, usually employed to disguise selfishness ("I'm doing it for me") as altruism ("I'm doing it for someone else").

⁴ Ziman, J. (1968) *Public knowledge: An essay concerning the social dimension of science*. Cambridge University Press.

⁵ For simplicity and historical precedent – and because most criminals are male – I use the masculine generic.

B. F. Skinner to psychoanalyst Karl Menninger⁶, provides a supposedly scientific basis for the “abuse excuse.” In the notorious Menendez case, for example (in an unprovoked attack, the Menendez brothers murdered both their wealthy parents), defense attorneys introduced testimony about a history of child abuse. Defense lawyers routinely contend that beating of the child may excuse murder by the adult⁷. The scientific assumption that behavior is causally determined has gradually led to a widespread feeling that the concept of personal responsibility is a holdover from an earlier and less sophisticated age, when people lacked the insight into the springs of human action now afforded by modern psychology and psychiatry.

This view does not withstand analysis. I will argue that determinism and responsibility are not contradictory but mutually supportive ideas. About human behavior we are not that much wiser than our ancestors. The advances in psychology, while interesting and even important (especially to psychologists) are in many respects little better than Penelope’s knitting: for every three steps forward, when we learn something new, we go one or two steps backward – and forget something old. The role of childhood experience in the adult behavior of any particular individual is largely a mystery and (I will argue) is in any case irrelevant to the innocence or guilt of the Menendez brothers and others like them.

The Argument Against Personal Responsibility

The argument against the concept of personal responsibility runs something like this. To hold a man responsible, he must in principle be able to have acted other than he did – to have the power of choice, to have “free will.” But human behavior is in fact determined; hence human beings cannot be held responsible.

There seem to be two meanings to “free will,” an objective meaning and a subjective meaning. The objective meaning has to do with predictability, and control. If behavior is completely predictable under all circumstances – like the position and velocity of moving bodies – it makes little sense to say that it could have been other than it was: “predictable” can be equated with “determined.” If we can predict in detail what someone will do, we may well conclude that his behavior is not free, no matter what he feels about it. By this criterion, human behavior is pretty determined; just ask any long-married couple how well they can predict each other’s behavior.

Determinism is also defined objectively in terms of *control*. If we can affect something in a predictable way by our own actions, then its behavior is, by definition, determined. Human behavior is notoriously affected by, *controlled*⁸, by outcomes. If I offer you a choice between two boxes, one containing \$1000 the other a piece of coal, I can confidently predict your behavior. Conversely, penalties that are both extreme and certain will likely have a deterrent effect. As we will see in a moment, this kind of determinism is not contradictory to the idea of responsibility, but essential to it.

The subjective meaning of free will has to do with how a person feels. No matter how well Joe Bloggs’ wife can predict his behavior, Joe himself probably feels perfectly free. In other words, subjective and objective meanings of free will need not, and in general do not, coincide. The feeling of freedom can perfectly well coexist with perfect predictability and perfect determinism⁹.

⁶ Menninger, K. (1968) *The crime of punishment*. New York: Viking Press; Skinner, B. F. (1971) *Beyond freedom and dignity*. New York: Bantam Books.

⁷ See Dershowitz, A. (1994) *The abuse excuse*. Boston: Little, Brown for an entertaining description of the creative flights of the legal profession in this direction. See Horowitz, D. L. (1986) Justification and excuse in the program of the criminal law. *Law and Contemporary Problems*, 49, 109-126, for an account of some of the legal aspects.

⁸ Skinner’s term – but see footnote 28.

⁹ Note that the contrary proposition, that we are, or could be, aware when our behavior is determined, leads to endless regress: because we are not then aware that our awareness may itself be determined. Note also that *failure* to

It is worth noting that introspection can reveal our behavior to be determined in at least the following sense: we may feel free to choose how we achieve our goals, but in general we do not feel free to choose the goals themselves. It is notoriously difficult to decide *not* to like tobacco or sweet things. The set of “reinforcers,” in behaviorist jargon, is largely predetermined. We may choose not to seek some goals, but that is only because they are subordinate to others (we decide not to smoke not because we cease to enjoy smoking, but because we want to avoid lung cancer). Usually, felt “free will” resides not in our choice of goals so much as in the means we choose to attain them.

Society and the law have no interest in subjective feelings (we are not yet in the era of George Orwell’s “thought police”). B. F. Skinner’s influential criticism of free will takes the objective view for granted: “In the traditional view, a person is free. He is autonomous in the sense that his behavior is uncaused. He can therefore be held responsible for what he does and justly punished if he offends. That view, together with its associated practices, must be re-examined when a scientific analysis reveals unsuspected controlling relations between behavior and environment.¹⁰” Thus, for Skinner, and most other psychologists, as well as many in the legal profession, determinism (“ unsuspected controlling relations”) = lack of free will = no responsibility for one’s actions. Or, to put the same thing in the opposite way: unpredictable, hence uncaused, behavior = free will = responsibility. If all behavior is determined, then, by definition, in any given circumstance, no man could have acted other than as he did. Hence, man does not have free will, hence he cannot be held accountable, he is not responsible.

What Does Responsibility Mean?

This is an odd argument, because it defines a responsible person in such a way that the usual sanctions (i.e., rewards and punishments) become ineffective, hence irrelevant. If the behavior of a responsible person were indeed “uncaused,” if such a man could in principle act in a completely unpredictable way, it is hard to see what purpose would be served by holding him accountable, that is, by delivering reward or punishment consequent on his actions. For example, suppose that a man were constructed in the following way, that at a choice point, when he might behave well or ill, he routinely decides at random, by tossing a mental “coin.” Sometimes he will behave well and sometimes badly; he has “free will”, as far as any external observer can tell – at least in the sense that his behavior cannot be predicted and has no (external) cause (i.e., it satisfies Skinner’s definition of free will). But it would make no more sense to hold such a man accountable than to blame the weather for spoiling your picnic. Why? Not because he is unpredictable – even if the weather were perfectly predictable, we still would not blame it for failing to fit in with our plans. The reason we would not blame such a man, I will argue, is because his behavior, like the weather, is *not responsive to contingencies of reinforcement*¹¹. It will make no difference, to “random man,” or to others like him (i.e., to what I will term his *deterrence group*), whether he is rewarded or punished. His future behavior will be the same. Consequently, a man who does not respond to reward or punishment cannot be held responsible. You might as well punish a paving stone for tripping you or a cloud for making you wet. I conclude that equating “responsible” with “possessing free will” with “uncaused behavior” is false. The behavior of a responsible man cannot be “uncaused” because he must be able to learn from the consequences of his behavior. I there-

predict does not imply *absence* of determinism. The behavior of a chaotic system like the logistic map becomes less and less predictable further and further into the future, even though it is perfectly deterministic, for example.

¹⁰ Skinner, B. F. (1971) *Beyond freedom and dignity*, p. 17.

¹¹ Dennett, D. *Elbow room*. Levin, M. (1997) *Why race matters*. New York: Praeger. Pp. 172 and 318 et seq. Staddon, J. (1995) On responsibility and punishment. *The Atlantic Monthly*, Feb., 88-94. I describe the concept of contingency in more detail in a moment.

fore propose that the first meaning for responsibility is *the ability to respond appropriately to reward and punishment*.

Ability to learn from consequences does not exhaust what we mean by “responsible,” however. The second meaning of responsibility is social. Responsibility is responsibility *to* something or somebody – to the state or to your family, for example. An orphan Robinson Crusoe without hope of rescue or future human encounter can be responsible to no one¹². More on the social aspect in a moment. I will argue that both aspects, sensitivity to reward and punishment and the social aspect, are essential to the legal concept of responsibility.

Uncertainty – a Red Herring. The physics of chaos and uncertainty are sometimes used to recapture free will and evade the conclusion that human behavior is determined¹³. The argument has particular appeal to physicists¹⁴, because chaos theory, Heisenberg’s uncertainty principle and quantum mechanics, all show that the idea of perfect predictability is wrong in physics. As I just tried to show, the evasion is unnecessary. The key point for free will and responsibility is not predictability but sensitivity to reward and punishment.

There is a paradox, however, that I will explore further in a moment. Sensitivity to reward and punishment depends on variability in behavior, which implies some degree of unpredictability. The paradox is that the predictability of behavior conferred by reinforcement learning (i.e., learning guided by reward and punishment) depends on initial *un*predictability. In trial-and-error learning, the organism must first behave more or less at random, but as the correct response is discovered, behavior becomes more predictable. Sensitivity to contingencies (see below) makes behavior predictable, but not (by definition) unalterable. Hence, an organism driven entirely by stimulus-response instincts cannot be held responsible – not because its behavior is predictable but because it cannot be changed by experience.

The Social Aspect

A dog, a lunatic and an infant are all sensitive to reward and punishment, yet we would not want to call them responsible beings. Fido may learn to stay off the Aubusson if he is consistently punished for doing so, but he lacks *something*. What is it?

Since the M’Naghten rule defining the insanity defense,¹⁵ it has been thought that though he be guilty of committing the *act*, a man cannot be found guilty of the *crime* if he did not know the act was wrong. But “knowing it is wrong” is only part of the story. “Knowing it is wrong” can be detected in an objective way *after* the act, in animals as well as humans. Many a dog owner can recognize the “guilty” behavior of his pet after it has transgressed. A criminal lunatic may know his behavior is wrong (or at least, punishable) and show it by trying to conceal evidence of the crime and perhaps showing remorse when caught. But neither dog nor man may be deterrable in the normal way. These examples show that “knowing it is wrong” does not by itself confer responsibility; the social component, the ability to be

¹² He may feel remorse for past misdeeds, but he is responsible to no one for his actions on the desert island.

¹³ In fact, for many practical matters, chaos theory is not relevant. Space shuttles and geostationary satellites achieve their orbits through Laplacean calculations. The “physics of the macroworld” as it is called *is* pretty much predictable. And we have no solid evidence that human behavior, at the gross level that is important for much public policy, is unpredictable either.

¹⁴ See, for example, Penrose, R. (1990) *The emperor’s new mind*. Oxford University Press.

¹⁵ In 1843, Daniel M’Naghten murdered Edward Drummond, secretary to Sir Robert Peel, the British Prime Minister who was M’Naghten’s intended victim. M’Naghten contended that his mission was guided by the “voice of God.” This trial yielded the M’Naghten Rule: insanity is proved if the defendant was “labouring under such a defect of reason, from disease of the mind, as not to know the nature and quality of the act he was doing; or if he did know it, that he did not know he was doing what was wrong.” See, for example, Livermore, L., & Meehl, P. (1967) The virtues of M’Naghten. *Minnesota Law Review*, 51, 789-856.

deterred, is also necessary. The issue here is not one of reward and punishment only, but of the ability to understand the contingency in advance. The dog will learn to stay off the rug, but we have no way to tell it in advance to stay off. Much the same holds for infants and lunatics. We may reward and punish them so that their behavior may be improved; but we do not expect them to learn from instruction or example. Punishing one dog will not improve the behavior of the canine community as a whole. Others cannot learn from the fate of the guilty. In a community that lacks the capacity to learn from instruction or example, punishment cannot *deter*.

In sum, responsibility has both an individual and a social component. The individual component has to do with the capacity to adapt to contingencies of reinforcement. The social component¹⁶ has to do with the capacity to be deterred by the prospect of punishment and encouraged by the anticipation of reward. Normal adult human beings possess both components; infants, many animals and some lunatics possess the individual component but lack the social; still other animals, and severely deranged human beings, lack both.

The Concept of Contingency

Reinforcement contingency is the behavioral psychologist's term for the relation between what an organism does and what, of consequence, happens to it. What must a man do to get the food, the money, or the girl? Reinforcement contingency is equivalent to the economist's equally technical term *objective function*.¹⁷ Rational-choice economists¹⁸ and naïve reinforcement theorists usually end the story there. For both, once the contingencies are set, the behavior is perfectly determined. If the hungry pigeon must peck a key ten times to get a bit of food (a fixed-ratio 10 schedule of reinforcement) it will inevitably learn to do so and go on to develop the characteristic stereotyped pattern of behavior. If single women are paid substantial amounts of child support, they will have more children and the rate of illegitimacy will increase.¹⁹

It is hard to overestimate the pervasiveness of the idea of reinforcement contingency in human social thought. It is, after all, far from a novel idea and it has led to practices that are generally very effective. Think, for example, of the behavioral technology involved in the workings of the great sailing ships, especially the great warships of the 18th and early 19th centuries. A single "ship of the line" might carry 1000 men, each a cog in a vast machine to work 100 guns and control dozens of sails in a coordinated way to attack an enemy or ride out a storm. Without habits of automatic obedience and mindless skill, such ships must soon have perished – and of course many did. But most did not, weathering storms, reefs and lee shores that all pose extraordinary risks for a cumbersome vessel entirely at the mercy of external forces. The rewards and often severe punishments that maintained this efficiency are well known and often deplored. Yet without them, the tight control necessary to the survival of large sailing ships alone for months on the high seas might well have been impossible. On a less heroic scale, the behavior of car drivers is quite sensitive to the penalties for motoring infractions. British drivers

¹⁶ I term this component *social*, even though it represents a capacity of individuals, because it depends on the uniquely human ability to learn about consequences through signals (signs, language) rather than only from direct experience.

¹⁷ See, for example, Staddon, J. E. R. (1979) Operant behavior as adaptation to constraint. *Journal of Experimental Psychology: General*, 108, 48-67, and other papers in that issue, for many examples of the relations between psychological and economic concepts.

¹⁸ The neoclassical economist, at least. There are other varieties that take a broader view (e.g., Arman Alchian, Brian Arthur, Ludwig von Mises).

¹⁹ It is of course irrelevant to the economic argument whether support payments actually reward childbearing or simply remove a disincentive. Charles Murray has provided a path-breaking analysis of this and related issues: Murray, C. (1984) *Losing ground: American social policy: 1905-1980*. New York: Basic Books.

seem to pay little attention to speed limits, for example, because penalties for speeding are rarely enforced. But drivers in those US states that stringently enforce speed limits are much more careful, even when the limit is well below what most would judge safe. All governments know the power of tax code to boost some activities (like house ownership, in the US and agriculture in the EU) and inhibit others (like alcohol and tobacco consumption).

Contingencies do not always work as intended, however. The failures are of two kinds: either the subject is smarter than the experimenter, or the reverse. Examples of the first kind are tax “loop-holes.” A provision designed to facilitate one kind of activity is turned to unintended uses by ingenious tax-avoiders. As the fraction of the national treasure taken in by governments has increased over the past century, and as the central state has assumed responsibility for ever wider segments of national life (from national defense to day-care), tax law has become a Darwinian contest in which laws are drafted for one set of purposes and then turned by taxpayers to quite others.

But contingencies may also fail because the subject is dumber, or more constrained, than the experimenter thinks he is. Pigeons (and even, alas, many human beings) cannot learn calculus, for example, and no matter how rewarding the schedule, a bird that must get its food only by forming derivatives will starve. So there are capacity limitations: pigs cannot fly and ants cannot speak. But there are also limitations of a more subtle kind. For example, raccoons can easily manipulate small objects and so many years ago some associates of B. F. Skinner’s, the Brelands,²⁰ who had gone into the animal-training business, tried to train a raccoon to put a wooden “coin” into a large piggy bank. They used the usual method of “shaping” – training by rewarding successive approximations to the target task. First reward touching the coin, then reward picking it up, then reward moving to the piggy bank, and finally reward insertion of the coin into the bank slot. In this way the raccoon eventually learned to get food reinforcement by putting the coin into the bank. But very soon the behavior disintegrated. Instead of putting the coin into the slot, the raccoon began to hold it in its paws and “wash” it in the characteristic way that these beasts do with small bits of food. By behaving thus with the token “food” the animals failed to make the correct, rewarded, response, hence blocked access to real food. The Brelands called this “irrational” misbehavior *instinctive drift* and discovered many other examples with other species. In every case, the “reinforced” behavior was superseded by an “instinctive” reaction that interfered with the behavior required by the reinforcement contingency and thus prevented delivery of the reinforcer.

Procedures that present painful stimuli (aversive schedules) provide many other examples. For example, if two rats, strangers to each other, are placed together in a cage and intermittently shocked through the floor they will attack each other and this will prevent them from learning to press a lever to prevent shock delivery. Even isolated rats find it difficult to learn an avoidance response because the shock induces instinctive reactions, such as “freezing”, that interfere with lever pressing. Most surprisingly, many animals can actually be trained to produce painful stimuli. After an appropriate training history, cats and monkeys will press a lever that occasionally produces painful electric shock²¹.

The failures of human aversive control are legion. For example, many of the British penal colonies in Australasia incorporated amazingly sophisticated systems of (usually aversive) behavioral control. Some worked satisfactorily, many did not. Norfolk Island, a tiny speck in the Pacific, provides exam-

²⁰ Breland, K. & Breland, M. (1961) The misbehavior of organisms. *American Psychologist*, 16, 661-4. See Staddon, J. E. R., & Simmelhag, V. (1971). The “superstition” experiment: A reexamination of its implications for the principles of adaptive behavior. *Psychological Review*, 78, 3-43, for a theoretical account of these data.

²¹ See Honig, W. K., & Staddon, J. E. R. (Eds.). (1977). *Handbook of operant behavior*. Englewood Cliffs, N.J.: Prentice-Hall, for many of these examples.

ples of both sorts. Under the regimen of Major Joseph Anderson (in charge from 1834 to 1839), for example, five men might get 1500 lashes between them before breakfast – for crimes like “singing a song” or “robbing the earth of its seed” (not planting corn properly). “Deranged by cruelty and misery, some men would opt for a lifetime at the bottom of the carceral heap by blinding themselves, thus, they reasoned, they would be left alone.”²² The only convict to leave a record of life on the island at this time wrote of Anderson’s even more brutal predecessor James Morisett:

“If you endeavour to take out of [a prisoner] that manly confidence which ought to be cherished in every civilized human being, you then begin the work of demoralization; and it will end in the very Dreggs of debasement & and insensibility to every species of integrity and decency, and eradicate every right feeling in the human breast. You make him regardless of himself, and fearless as to the consequences of doing wrong to others”²³.

Clearly, aversive contingencies can fail if they are too severe. The reason? The “misbehavior” effects mentioned earlier. As in the rat example, punishment that is excessive produces behavior that competes with the behavior the punisher desires, and thwarts the intended effect of the punishment regimen.²⁴ Punishment that is perceived as unjust also elicits reactions (usually aggressive) other than those intended.

It is noteworthy that Norfolk Island was also the site of a much more moderate regime that resembles in almost every particular the modern behavioral psychologist’s *token economy*, still widely used in such mental and remedial institutions as remain after the closures of US institutions in the 1970s and ‘80s. These ideas were put in place on Norfolk Island between 1840 and 1844 by one Alexander Maconochie: “Let us offer our prisoners, not favors, but *rights*, on fixed and unalterable conditions.”²⁵ Called the “Mark System,” Maconochie’s scheme involved sophisticated scheduling of both positive and negative reinforcement.²⁶ Good behavior was rewarded with “marks” that went towards a reduction of sentence: 6000 marks were equivalent to a seven-year sentence, 7000 to 10 and 10 thousand to life – a nonlinear scale corresponding roughly to the disutility of each sentence. Marks could also be exchanged for goods, they were “just wages, and will equally stimulate to care, exertion, economy and fidelity.”²⁷ By all accounts, the Mark System was reasonably successful during its short life, but it was in the end stymied by Britain’s need for Norfolk Island to be a place where “Felons...have forfeited all claim to protection of law”, the deterrent *ne plus ultra* for intractable villains. Maconochie’s humane system failed of *this* objective.

Contingency vs. Cause

²² Hughes, Robert (1987) *The fatal shore*. London: Collins, Harvill. P. 481.

²³ Frayne, Laurence *Memoir on Norfolk Island*. NSW Colonial Secretary Papers, vol. 1 (re. NSW 1799-1830). Ms. 681/1, ML, Sydney. Pp. 25-26.

²⁴ Literature provides many vivid examples where the threat of severe punishment fails to achieve the desired behavior. For example, in the 17th century play *Surgeon of Honor*, by Spanish writer Calderon de Barca, the clown-servant Coquin is given the ultimatum that he must make the King laugh – else have all his teeth pulled out. This particular reinforcement contingency is more likely to leave Coquin toothless than the King in fits of mirth.

²⁵ Hughes (1987), p. 499.

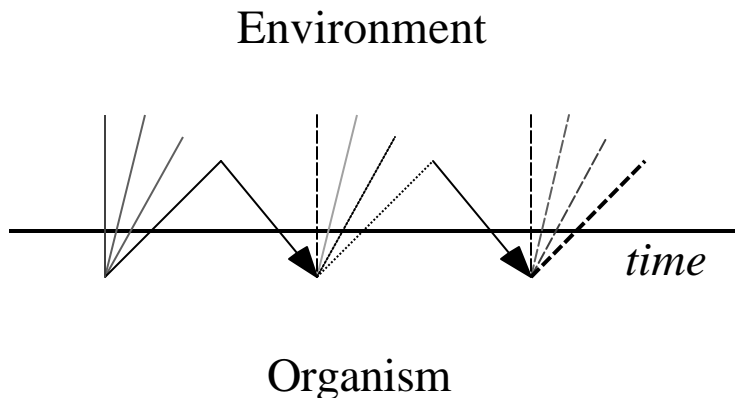
²⁶ Negative reinforcement is the removal of an aversive stimulus (here, the convict’s time of servitude); positive reinforcement is the goods that could be bought with “marks.” “Time off for good behavior” is the modern version of Maconochie’s “marks.”

²⁷ *Ibid.*, p. 500.

So, despite its many successes, we must conclude that reinforcement contingencies by themselves do not always allow us to predict²⁸ behavior. Why? The simple philosophical reason is that explanation in terms of contingency is *teleological*²⁹. To explain behavior by the maximization of utility (the economic idea of *rationality*) – or in terms of its consequences in general – is to explain it in terms of its ultimate goal not its proximate causes. To say that men act rationally (should they ever do so), is not to say what rationality *is*. As is well known, means are always finite; hence behavior is not always optimal.³⁰ Economists have proposed various fixes for optimality analysis, such as Simon’s “bounded rationality”³¹ (the idea that people will settle for results that are “good enough” not necessarily “the best”) and others. But such is the richness of human behavior, none is really adequate. Indeed, in the long run, the only satisfactory solution is to abandon the whole idea of contingency as an explanation and look directly for the proximal causes, the mechanisms, of behavior.

We are a very long way from that point. But there is an intermediate approach that preserves the insight that behavior is usually guided by contingencies, but can prevent us from falling into the trap of teleological determinism. I have called this the *Darwinian metaphor*³².

Contingencies reward and punish, but they can only reward and punish *what has actually occurred*. In other words, they can only *select* from what is offered by the subject. Moreover, rewards and punishments have two kinds of effects: they select, yes, but they also have what I will call *direct effects*³³ – shock-elicited fighting and freezing, for example, “washing” of the food-associated tokens and others. Problems arise if the direct effects of rewards and punishments conflict with their selective effects.



The general process of learning through contingencies of reinforcement (operant learning) is illustrated schematically in the figure. The fan of lines indicates the *repertoire* of behavior at a given time. The particular response that has some (re-

²⁸ “Control” in Skinner’s misleadingly forceful terminology. The effects of contingencies are always limited by the means organisms have for adapting to them; hence, “predict” is a more accurate term than “control.”

²⁹ More precisely, *teleonomic*, in Pittendrigh’s sense (Pittendrigh, C. S. (1958) *Adaptation, natural selection, and behavior*. In A. Roe & G. G. Simpson (Eds.) *Behavior and evolution*. New Haven: Yale University Press.

³⁰ There are extensive discussions of these issues in the biological literature. See for example, Maynard Smith, J. (1978) Optimization theory in evolution. *Annual Review of Ecology and Systematics* 9:31-56. Oster, J. F., & Wilson E. O. 1978 A critique of optimization theory in evolution. In *Caste and ecology in the social insects*. Princeton University Press, 292-315. Staddon, J. E. R. (1983). *Adaptive behavior and learning*. New York: Cambridge University Press. Staddon, J. E. R. (1987) Optimality theory and behavior. In J. Dupré (Ed.) *The latest on the best: Essays on evolution and optimality*. Cambridge, MA: Bradford/MIT Press. Pp. 179-198. Staddon, J. (1993) *Behaviorism: Mind, Mechanism and Society*. London: Duckworth. Staddon, J. E. R., & Hinson, J. M. (1983). Optimization: A result or a mechanism? *Science*, 221, 976-7.

³¹ Simon, H. A. (1987/1978) Rational decision making in business organizations. In *Advances in behavioral economics. Vol. 1*. L. Green & J. H. Kagel (eds.) Norwood, NJ: Ablex Publishing. (The 1978 Nobel lecture).

³² Staddon, 1993.

³³ These direct effects need not be *immediate*. Instinctive drift follows the laws of Pavlovian (rather than operant) conditioning and takes some time to develop. The point is that what I am calling direct effects are effects of antecedent, not consequent, events. Psychologists sometimes use the term *elicited* for these effects.

inforcing) consequence ends in a black arrow, which leads from the behavior back to the organism, symbolizing the effects of the consequence on the repertoire in the next instant of time. Notice that the set of rays is a little different after the first “reinforcement”. The process continues and, if the contingencies are working effectively, the repertoire will eventually be dominated by the desired, “target” response.

But always it is the organism that “proposes” and the environment that “disposes.” If, as sometimes happens, the repertoire elicited by the reinforcer fails to contain the desired behavior, training will fail – as it so often did on Norfolk Island, where the extremity of punishment produced despair and resentment, rather than productive action. The lesson? We must take as much account of the direct effects of our reinforcers as of their selective effects. The error of rational-choice economists and radical behaviorists alike was to ignore the direct effects almost entirely.³⁴

Responsibility, Punishment and the Authentic Self

In the world of media politics “accepting responsibility” often means nothing more than saying “I accept responsibility,” but in the real world it means paying a cost or accepting a punishment for failing in some obligation. The obvious policy question, therefore, is, How efficacious *is* punishment, both individually and as a deterrent?

Psychologists and psychiatrists, both behavioristic and otherwise, have answered (almost³⁵) with one voice: punishment (“aversive control” in Skinner’s terminology) is not effective. The reasons vary. The touchy-feely types, for example, argue like this: “any of us who were raised in the traditional patriarchal system have trouble relating because we’ve been ‘mystified’ to some degree by an upbringing that compels obedience and rules by fear, a raising that can be survived only by a denial of the authentic self.”³⁶ What this kind of New-Age vapping really means is far from clear; but the echoes of Jean-Jacques Rousseau are plangent. The general ideas seem to be: (a) Each of us has an “authentic self,” akin to J-J’s “noble savage”; (b) the authentic self is intrinsically free and good; but is (c) corrupted by conventional authority – the “patriarchal system,” “rule by fear,” etc. (“Man is born free, but is everywhere in chains” and so forth.)

Each of these notions is doubtful. Proposition (a) assumes that each of us has an identifiable essence. No proof is offered, perhaps because none is possible. What difference would it make if there were no “authentic self”? How can such a self be identified? Does a mass-murderer like Charles Dahmer have an authentic self? Is that self a mass murderer or something better? If the former, can we hold up the authentic self as an ideal? If the latter, how do we know? Which brings us to Proposition (b): supposing there is an authentic self, how sure can we be that it is in every case socially desirable? Answer: We cannot. *Lord of the Flies* is at least as likely to be an accurate picture of man’s true nature as *Peter Pan*. So to (c), where the evidence is also contrary to Rousseau’s view. Nothing is more self-centered, amoral and uncaring than a baby. Civilization depends upon inculcating, not always by positive reinforcement, altruism and conscience in little beings that begin with few vestiges of either. (The process was imperfectly successful in the case of J-J himself, as we know, since he forced his illit-

³⁴ This actually gives too much credit to the economists, who were entirely uninterested in dynamics, process, etc.

³⁵ The work of Sarnoff Mednick, who tracked the careers of criminals with and without a history of consistent punishment for their crimes, is a notable exception. For these individuals, punishment was evidently a strong deterrent; e.g., Brennan, P. A; Mednick, S. A. (1994) Learning theory approach to the deterrence of criminal recidivism. *Journal of Abnormal Psychology*, 103, 430-440.

³⁶ John Bradshaw (1993) *Lear's* magazine, January, p. 42.

erate common-law wife to put all their offspring into foundling homes so they should not distract her from devoting full attention to *him*.)

So much for pop psychology: what thinks the mental-health profession? There are strong economic reasons for professional psychotherapists to downplay aversive control. Hence, it is no surprise to see little support for it within the profession. “Clients” usually seek psychotherapy voluntarily. Indeed, the therapist must live by his ability to attract them. There must be rewards for undergoing psychotherapy. If therapy were generally effective the reward would be obvious: a cure.³⁷ But the evidence in favor of “the talking cure” is in fact tenuous. The best that can be said is that although average mental health (however measured) after therapy is much the same as before, the variance has increased. That is to say, some patients get better but an equal number get worse.³⁸ Lacking the ability to cure reliably, psychotherapists must give their patients some other reinforcer to keep them coming. Generally speaking, a regimen that emphasizes punishment and discipline³⁹ and will fail to please. Since no profession can for long embrace beliefs that threaten its own existence, it is natural that hostility to aversive control has come to be received wisdom among psychotherapists.

It’s also worth noting that the professional psychotherapist has an obligation *only to his patient*. The impact of his work and his ideology on others and on society at large is not part of his role. If, as I will argue, judicial punishment is most important for its social effects, we should not be surprised to find little acknowledgment of this fact in the mental-health profession.⁴⁰

Skinner and his school argued against punishment for another reason. Like Sigmund Freud and Karl Menninger, Skinner was a master at waving the banner of “science” whenever his arguments began to violate common sense. He did not reject punishment because it is morally wrong, which would be a defensible, if controversial, view. Like all behavioral scientists, he believed he had rigorously excluded morality from his work. Instead, he defended an ostensibly absurd position by asserting that “science shows” it to be true. He argued that aversive control must be rejected because it doesn’t work.

Skinner’s argument boils down to two points: punishment is ineffective because when you stop punishing, the punished behavior returns; and punishment provokes “counterattack”. Both have proved to be either wrong or to apply with equal force to positive reinforcement. The correct conclusion is that evidence from the experimental laboratory is pretty neutral in deciding between reward and punishment.

³⁷ Seventeenth-century diarist Samuel Pepys was willing to be “cut of the stone”, *sans* anesthetic, because without the operation, death was certain. Depressives will tolerate the pain of electroconvulsive shock because the treatment, crude as it is, often provides relief. People will accept pain from a therapist if their condition is dire and relief is probable.

³⁸ A critical discussion of psychotherapy, particularly psychoanalysis, that is both knowledgeable and entertaining is experimental psychologist Stuart Sutherland’s (1976) description of his own treatment for mental-health problems: *Breakdown: A personal crisis and a medical dilemma*. London: Weidenfeld & Nicolson. A more recent critique is: Terence W. Campbell (1995) *Beware the talking cure: psychotherapy may be hazardous to your mental health*. Sirs, Inc. It is not just capitalist insensitivity that makes the bean-counters who run health maintenance organizations reluctant to pay for mental-health benefits. The cost-benefit calculations are not favorable.

³⁹ As induction into some religious orders – a sort of psychotherapy, one might suppose – used to do.

⁴⁰ This myopia seems to be characteristic of all the “caring” professions. Some readers may recall the objections of physician Dr. Marcia Angell, editor of the *New England Journal of Medicine*, to a government-funded study that proposed to compare very low doses of anti-HIV drugs with a no-dose control group, as part of an effort to find an inexpensive treatment for endemic AIDS in sub-Saharan Africa (See Marcia Angell, *Tuskegee Revisited*, *Wall Street Journal*, October 28, 1997). Her objection: that some experimental subjects received no treatment. The fact that no subjects were injured, that the value of the “treatment” given to experimental subjects was conjectural, and that the results of the study might benefit millions, carried no weight. Her attention was entirely focused on the fact that the control subjects (under the “care” of the experimenters, in her view) received no treatment. (See Staddon, J. *Wall Street Journal*, letters, November 11, 1997, for a contrary view.)

They both have their advantages and disadvantages: punishment is better for suppressing behavior, positive reinforcement better for generating behavior; some aversive schedules have more persistent effects than any schedule of positive reinforcement,⁴¹ and so on. There is no scientific argument for the abolition of punishment.

⁴¹ In the animal laboratory, the most persistent behavior is generated by *shock-postponement* schedules. These work by delivering brief, painful electric shocks at fixed intervals. The next shock can be postponed for a fixed time by pressing a lever or making some other operant response. Well-trained animals respond frequently enough to avoid all shocks, and may continue to do so indefinitely, long after the shock generator has been disconnected. No schedule of positive reinforcement produces such persistent effects. (It is an interesting sidelight on the nexus between policy advice and laboratory science, that the inventor of this procedure is nevertheless a passionate opponent of aversive control in public policy: see Murray Sidman *Coercion and its fallout*. Boston, MA: Authors Cooperative, 1989.)

Judicial Punishment

If it were not for the Gallows, some men are of so cruel a nature, as to take delight in killing men - more than I should to kill a bird. Thomas Hobbes⁴²

Legal scholars normally identify two purposes for judicial punishment, retribution and deterrence.⁴³ These two ideas are not in fact as distinct as they sound, because retribution, the public sense that justice has been done, itself constitutes a deterrent to potential malefactors. So I will assume that deterrence is the chief function of judicial punishment⁴⁴. Does punishment really deter? The argument is clouded by ideology – many people are ethically opposed even to just punishment, and most people are opposed to “cruel and unusual” punishment. And it is in fact strictly impossible to decide the issue *conclusively* by the methods of science. We cannot do experiments on human society, so arguments about the deterrent effect of (for example) the death penalty must always be inconclusive. Did the murder rate decrease in State X when the death penalty was imposed? If the answer is “yes” the objection by death-penalty opponents may well be that the penalty was imposed only because the murder rate at the time was unusually high – and would be expected to fall anyway (because of “regression to the mean”), so that fall was not caused by the penalty. But if the answer was “no” (the murder rate increased) the objection by death-penalty proponents will be that the rate would have been even higher without the penalty – or that an increased penalty takes time to have an effect (“Let’s look five years down the road...”), and so on. These arguments cannot be ruled out. The problem is that in the real world murder rates are affected by many things in addition to the penalties for murder. And in any case the causation goes both ways: penalties affect crime rates, and crime rates affect penalties. I propose that we just assume *some* deterrent effect of judicial punishment, and then ask how it should be sensibly applied. We can then come up with a very simple, utilitarian, view of personal responsibility.

The utilitarian view is that the purpose of legal punishment is to minimize the total amount of suffering in society, the suffering caused by crime as well as the suffering caused by punishment. The concept is simple: if theft is punished by amputation, the level of theft will be low, but the suffering of thieves will be very high, higher perhaps than warranted by the reduction in theft.⁴⁵ On the other hand, if murderers go free, the level of murder will be high and the pain of the innocent will hardly be compensated by the euphoria of the killers. We may argue about how to measure suffering and how to assess the

⁴² In John Aubrey’s *The life of Mr Thomas Hobbes of Malmesburie*, in *English biography in the seventeenth century*. Vivian de Sola Pinto (Ed.) London: Harrap, 1951. P. 188.

⁴³ The commonest form of punishment, incarceration, serves a third purpose: prevention of further offenses by the criminal during his prison time. Incarceration of repeat offenders can be very effective in reducing crime rates (e.g., Methvin, E. H., 1997, Mugged by reality. *Policy Review*, July, 32-38). But the control here is physical not behavioral.

⁴⁴ Many legal theorists have argued for the contrary view. Most recently, Michael Moore (see *Placing blame: A general theory of the criminal law*. Oxford: The Clarendon Press, 1997) has argued persuasively and at length that “Of the possible functions for criminal law, only the achievement of retributive justice is its actual function.” (pp. 78-9). I have no space here to contest this position. Suffice it to say that (a) achieving consensus on what *is* retributively just is likely to be even more difficult than getting agreement on what minimizes general suffering. (b) The near-universal agreement that punishment for attempted murder should be less than for successful murder strongly supports the idea that judicial punishment should have a retributive component and not be deterrent only. But, (c) punishment certainly *does* deter, so it seems unwise to formulate criminal justice policy without taking this element into account.

⁴⁵ Philosophical economists have pointed out that utility cannot be compared between individuals: even President Clinton cannot really “feel my pain.” Nevertheless, such comparisons are essential to any utilitarian analysis of social effects. In practice, most writers seem to assume a highly nonlinear scale that effectively rules out extreme (“cruel and unusual”) punishments. John Rawls’ influential *A Theory of Justice* (1971. Cambridge, MA: Harvard University Press) also argues for this view.

effect of a given level of legal punishment for a given crime, but the utilitarian view⁴⁶ of punishment, seems reasonable enough. It is consistent with the fundamental principle that government exists not to empathize with individuals (although empathy is integral to the political process), but to act *for the welfare of society as a whole*⁴⁷.

In the utilitarian view, the case for *individual* responsibility rests entirely on the beneficial *collective* effects (on the sum total of human suffering) of just punishment. It does not rest on philosophical notions of individual autonomy, or on personal intent, or anything else at the level of the individual – other than normal sensitivity to reward and punishment. The idea that the law is somehow concerned with the mental state of the accused, rather than with his susceptibility – and the susceptibility of others like him (his deterrence group) – to judicial sanctions, is a recent error which is aided and abetted by the myopic individualism of the mental-health profession.

Surprisingly, perhaps, traditional legal practice agrees with the utilitarian view. Only when punishment is likely to be completely *ineffective* as a deterrent does the law limit its use. If a criminal is insane – which is to say that he is not sensitive to rewards and punishments in the ways I described earlier – he is not found guilty. If a prophetic angel had whispered in the ear of Lyle Menendez “If you shoot your mother, you will die” and if Lyle had believed the angel – but shot anyway, then he is *not guilty*. (Although I would hope we would lock him up anyway, as dangerously insane.)⁴⁸ But if, in the jury’s estimation, he would have heeded such an angel, then he is *guilty*, because he was indeed responsible for his action in the only way that makes judicial sense.⁴⁹ More on this topic in a moment.

“Not guilty” is also the verdict if injury is the unintended result of an action that a reasonable person would expect to be harmless: if the politician’s wife turns on a light that then gives an assassin a clearer view, the law would not implicate her in his death – because, not knowing the risk, she could not have been deterred by legal penalties for putting someone at risk. “Guilt” is established not so much by the act itself, as by the potential of punishment to deter the act. Notice that this definition skirts the difficult issue of “motive” even as it comes to a similar conclusion, in most cases (a maid, complicit in the assassination plot, might be deterred under such circumstances, because her knowledge of the risk makes her liable). Traditional law is objective in a way that recent psychotherapeutically inspired distortions (e.g., the *Durham* decision, see below) are not.

What About Rehabilitation?

The utilitarian view of punishment and responsibility leaves the criminal largely out of the picture; he figures only as one among a multitude. The focus is on society at large. But what *about* the poor

⁴⁶ See Cesare Beccaria *An Essay on Crimes and Punishments* (1764, 1963), for what seems to be the earliest statement of this view.

⁴⁷ Why worry about the suffering of the criminal? some might ask. Why not just minimize the suffering of victims? One reason is that just retribution probably cannot be excluded as a component in criminal justice, even if it is probably not the only component. Another is that without including the suffering of the criminal in the utilitarian equation, we can raise no principled objection to a colleague of mine who, only half-joking, proposed the death penalty for littering.

⁴⁸ What about the individual who is sensitive to punishment in the normal way, but is willing to disobey the law anyway because he disagrees with it? Mohandas Gandhi and Nelson Mandela, but also Oliver North and Leon Trotsky, come to mind. Such individuals will naturally fail the deterrence test. Consistency implies that they should therefore be treated as irresponsible, but dangerous, like a mentally ill person. And indeed, political prisoners are often treated in this fashion, restrained indefinitely but not necessarily punished in any other way (cf. Mandela’s long-term incarceration, the indefinite house-arrest of Aung San Suu Kyi, etc.).

⁴⁹ A skeptic might ask, “Well, if he would have heeded the Angel, why did he shoot?” The obvious answer is “Because he thought he wouldn’t be caught – or, if caught, could get off.” The belief that a crime will go undetected or unpunished is of course a major “root cause” of crime.

felon, his suffering and his future? For much of this century, with rising public interest in both mental health and redemptive religion, the fate of the criminal has often been at center stage. In this, as in so many other things, Ann Landers⁵⁰ is a touchstone for popular opinion. She recently cited with approval the work of influential psychiatrist Karl Menninger as the last word on the subject of judicial punishment. What were his views? In a word: incredible.⁵¹ Menninger, more concerned with the criminal than the victim, called punishment a crime, argued that criminals are a creation of the law-abiding (who “need” someone to punish!) and believed criminality to be a form of mental illness. To Menninger, criminals should be *treated* – rehabilitated – not punished. He would not have been sympathetic to any view that legitimates punishment.

This is in some respects an understandable position. After all, the victim in murder cases is not present and cannot testify, whereas the accused, neatly groomed, remorseful and perhaps even subject to recent religious conversion, plucks at the heart-strings of the *simpatico* juror. Many are therefore drawn to Menninger’s apparently humane position. What are we to say to them? Just this: the issue is simply a matter of numbers. Unfortunately, like most social and political issues, we don’t know the numbers and cannot find them with certainty. But we can at least understand what it is we need to know. Granted that our aim is to minimize total suffering, we need to know (a) What is the probability that a typical criminal can be rehabilitated? (b) How much will he be improved by treatment, compared to his previous state: that is, how many crimes will he commit after a “cure” vs. how many would he have committed without? (c) How much effort (people, facilities) must society devote to this effort? (d) Could these resources be better employed elsewhere (i.e., what is the opportunity cost)? But the critical question is the last one: (e) How many potential crimes by others would be deterred by punishing this criminal rather than rehabilitating him (you may recall that this was the question that killed Maconochie’s reforms on Norfolk Island)?

The best answers to the first four questions seem to be (a) low, especially if he has a history of recidivism; (b) not much; and (c) a lot, and (d) almost certainly yes – all with a high degree of uncertainty. But the answer to the last question, with somewhat less uncertainty, is “several.” That is, by punishing one criminal for one crime, we are likely to prevent several others. Even if our criminal is not rehabilitated at all, therefore, punishment may be a very good deal for society. And, anything that weakens its deterrent effect, including even a very successful effort at rehabilitation, may be a bad deal. Without just punishment, the level of suffering in society may well be higher than with it. Dr. Menninger’s view is not so much soft-hearted as soft-headed.

Diminished Responsibility

I have discussed some cases of diminished responsibility – children, the insane – that can be understood simply. But there are others that require more analysis. How is responsibility affected by drugs? By a mental-health “disorder”? By “false consciousness”? In some cases, scientific analysis suggests concrete solutions, but in others analysis can only reduce the question to issues that are at bottom political and moral, not scientific.

Drugs and Responsibility

Not all drugs impair behavior: tobacco, for example, is addictive but does not in any way diminish responsibility. We can neglect drugs without effect on behavior. We need to distinguish between behavior-altering drugs that are addictive and those that are not. Alcohol, for example, is not addictive for most people, but does impair judgment. The alcoholic may, in his drunken state, be responsive to

⁵⁰ Ms. Landers, now in her eighties, has been writing an advice column syndicated in hundreds of U. S. newspapers for more than forty years.

⁵¹ Menninger op. cit.

reward and punishment, as the infant or the idiot are, but he clearly lacks the social component I referred to earlier. The drunk cannot make an accurate assessment of the consequences of his actions. *In his drunken state* his responsibility is diminished. But, someone not addicted to alcohol can both understand the connection between drunkenness and diminished responsibility and can be deterred by suitable contingencies from drinking alcohol at inappropriate times – before driving, for example.

For the alcohol addict, it is otherwise. He cannot be deterred from drinking in the same way as the non-addict. He may continue to drink inappropriately even in the face of extreme penalties. He is, in fact, not fully responsible. What should be done? First there is the problem of distinguishing between addicts and non-addicts. Is addiction all-or-none, for example? If it is, the problem is simpler, providing we have a reliable litmus test. If it is not, if addiction is a matter of degree, then the possibility arises that the number of “addicts” is dependent on the level of punishment. If punishment is severe, there will be fewer addicts than if it is lenient. But suppose addiction is all-or-none, and addicts can be reliably identified. Should alcohol be kept from them entirely, and if so, how? Since we cannot predict in advance what the general consequences of different policies will be, and since all of them are likely to conflict with some widely held moral beliefs (such as the intrinsic value of personal freedom, for example), these are political and moral questions not to be settled here.

And finally, what of the individual who has never tasted alcohol, who is neither addict nor social drinker? How should he be treated? Clearly, given the right contingencies, he can be deterred from sampling an unknown delight. But the penalties may need to be severe to deter a majority of potential drinkers.⁵² Because we (so far) cannot tell in advance who the potential addicts are, many individuals will be severely punished who would not go on to become addicts and would not, therefore, pose a threat to others. Is it fair to punish severely individuals engaged in behavior that *for them* is not potentially dangerous? Is it even fair to punish potential addicts, given that their addictive potential is not their fault? Once again, we are in the realm of politics and moral philosophy, questions go beyond the ambitions of this paper.

Psychological Disorder

Drug addiction may or may not be all-or-none, but mental disorders obviously come in a myriad varieties, from the clearly insane and incompetent to Woody Allen and friends who, in past ages, might have been questioned as to their character but not their health. Now all are subsumed under the heading of “mental health problems.” Unfortunately, the classification of mental disorders is not a science. The American Psychiatric Association's *Diagnostic and Statistical Manual of Mental Disorders* (DSM), the bible for mental-health professionals, is a strange mixture of science, fashion, politics and the economics of the medical profession.⁵³ In the words of a *sympathetic* commentator “[N]ew disorders do not represent claims to fresh biological knowledge. Which is to say that yes, a particular mental disorder is whatever the DSM says it is.”⁵⁴

DSM “disorders” may not represent biological knowledge, but all too often they do represent arguments for diminished responsibility. “The Devil made me do it” is no longer acceptable. But syn-

⁵² Of course, prohibition is one solution to this dilemma. But again, non-addicts are punished needlessly by being deprived of a harmless (for them) enjoyment. Is the greater good served? What about personal freedom, which is also a good? Again, these questions go beyond practical science.

⁵³ See, for example, the recent exposé *Making us crazy: DSM—the psychiatric bible and the creation of mental disorders*. By Herb Kutchins and Stuart A. Kirk. New York: The Free Press; and *The Encyclopedia of Insanity - A Psychiatric Handbook Lists a Madness for Everyone*. *Harpers Magazine* February 1997. By L.J. Davis.

⁵⁴ Diagnosis: Totally Sane. The DSM isn't crazy in the slightest. A review of Kutchins and Kirk by Larissa MacFarquhar in the Internet magazine *Slate*. Posted Wednesday, Nov. 12, 1997.

dromes from PMS to “road rage” are now routinely offered as mitigating circumstances in criminal trials. When does a mental disorder diminish responsibility, and what should be done about it?

Once again, normal susceptibility to the effects of reward and punishment suggests an answer. If the individual is so disordered that he is not deterred by the prospect of punishment, then his responsibility is diminished. But (it may be objected) surely we have merely shifted the problem. What is “normal susceptibility”? What is “normal punishment”? Even a very disturbed individual might well be deterred by a the prospect of a severe punishment, severe enough to be unacceptable to most people. We have in fact already encountered this problem, in connection with drug addiction. But in the case of mental disorders we know that there are grades of susceptibility, of “self control,”⁵⁵ in an older terminology. Indeed, quite apart from supposed mental illness, we know that there are individual differences in susceptibility to punishment, as there are in everything else. So the question of mental illness boils down to this: How are we to deal with individual differences in “self control”?

We can distinguish three cases, differing in the way that individual differences are identified, and the accuracy with which people can be categorized:

Self-identification. Most straightforward is the case of the repeat offender. An individual who breaks the law repeatedly, and is caught and punished, is clearly less susceptible to deterrence than average. And the law accommodates this difference by increasing the punishment for repeated offenses. This method is self-evidently valid and is not controversial. But notice that the reduced sensitivity to contingencies demonstrated by recidivism is not taken as a *mitigating* factor, but as a reason to increase the level of punishment. The opposite is true for mental disorders, which are always proposed as mitigating factors.⁵⁶

Identification by Behavioral Test. It is conceivable that a psychological test, like an intelligence test, could be devised to identify people less likely than average to respond to normal deterrents. Perhaps some of the disorders identified by DSM fall into this category. But DSM has other purposes and only the most extreme diagnoses, “paranoid schizophrenia,” “dementia,” etc. – those which place the individual in the conventional category of “crazy” – are generally agreed to mitigate criminal guilt.

But even if tests for milder degrees of insensitivity could be devised, it is by no means clear what should be done with them. Take, for example, the case of the hyperactive child, so-called “attention deficit hyperactivity disorder” (ADHD⁵⁷). Such children are often identified because they present discipline problems. In the past, schools used more severe punishments than are permitted now. And punishments were applied not just to misbehavior, but also for failure to learn. Perhaps this move towards a kinder, gentler schoolroom is a change for the better, but one certain result is that less amenable children may now misbehave or fail in a way that they would not have under stricter discipline. For some (how many?) children, the ADHD diagnosis has in effect shifted a self-identified insensitivity to normal rein-

⁵⁵ There is a more technical meaning for self-control, as capacity to delay gratification, which is treated at length in an elegant experimental literature. I intend the simpler meaning – susceptibility to deterrence – here.

⁵⁶ There is a category of diminished responsibility that is the mirror image of the repeat offender. The repeat offender is by definition insufficiently deterred by punishment. The complementary category is people who are excessively encouraged by reward. It may seem odd to see such people as less responsible than average, but this is apparently the view of a colleague of mine who recently termed payments to a public lottery an “implicit tax.” Granted that a tax is “a contribution exacted by the state: anything imposed, exacted or burdensome” to call voluntary and presumably pleasurable participation in a lottery a tax is on its face absurd – unless you feel that people who play a lottery cannot control themselves – are not fully responsible. This category is not yet fully accepted in the legal arena (“compulsive gambling” and “sex addiction” as mental illnesses are contenders, I suppose), but its time will probably come.

⁵⁷ See for example, Attention-deficit/hyperactivity disorder. By M. D. Rapport In *Handbook of psychological treatment protocols for children and adolescents*. Edited by V. B. Van Hasselt and M. Herson. Mahwah, NJ: Lawrence Erlbaum, 1998, (pp. 65-107) for the current view on this fluid category.

forcement contingencies from the “repeat-offender” category – which justifies upping the level of punishment – to the “mental illness” category, which justifies medication.

It is easy to see why society is suspicious of punishment of children. The practice is obviously open to abuses, even if the disciplinarians are entirely without malice. Consider two hypothetical children. Child A is simply at the “insensitive” end of the normal distribution of sensitivities, but Child B has real mental problems. The problem is that the two children can often be distinguished only by the fact that Child A does eventually learn to behave, whereas Child B does not, even when punishments have reached unacceptable levels.⁵⁸ In other words, by the time one can tell the difference, Child B has received cruel levels of punishment. The problem is where to draw the line. It is not to be drawn at “zero,” by eliminating punishment altogether (which is the aim of many educators) because the educational cost to the class as a whole outweighs the cost of mild punishment of miscreants.⁵⁹

The ADHD diagnosis is now used to identify Child B without raising his level of punishment to unacceptable levels. But it was not designed for this purpose. The bias in the mental-health profession against punishment, and feeble discipline in most public schools, means that the diagnosis is rarely tested against the obvious alternative – that the kid just needs tougher discipline. Hence, it is probably applied far too widely. Perhaps psychologists in the future can come up with valid tests not just for “attention deficit” but for sensitivity to reasonable discipline.⁶⁰

There is a strange variant of the psychological-syndrome justification that occasionally pops up in the literature dealing with personal responsibility, and that is a version of the Marxist notion of “false consciousness.” I encountered this idea most recently in discussions with a radical feminist who insisted that some 25% of college women are victims of rape – even though most questionnaires have come up with much smaller numbers. It turned out that she had a much broader version of rape than I. It included (for example) consensual sex while both parties are drunk. Most would assume that responsibility in such cases is shared, if not exactly equally, between the man and the woman. But the argument is offered that because women have been oppressed in U.S. society, i.e., conditioned in certain ways, they are likely to say “yes” when they mean “no,”⁶¹ and to feel coerced by their boyfriends, even in the absence of physical threat. Often, “victims” see things more clearly in the morning – or after spending a few sessions with a support group⁶² – when they realize that they were not really responsible for what happened. So, they must have been raped.

Responding to an apparent rise in spousal abuse, the legal systems in some states seem to echo the view that women have diminished responsibility. For example, on being notified of a “domestic dispute” where physical violence is *possible*, police are often obliged to arrest the man, even if the woman protests, whether or not there is evidence of coercion or intimidation. The woman is assumed to be not responsible, as if she were a child or mental defective.⁶³ When may the law legitimately treat a normal

⁵⁸ There are of course many more dimensions to the problem than simply punishment and reward. The child is sensitive to all sorts of aspects that cannot easily be quantified: Is the punishment just? Is the teacher to be respected? To be believed? What is right and wrong? Will my parents support me or my teacher? How about my friends? And so forth. These effects, which may deflect the expected effects of reward and punishment, come under the heading of the direct effects of reinforcement contingencies, discussed earlier.

⁵⁹ See Charles Murray, *op. cit.* for a thoughtful discussion of this problem.

⁶⁰ “Repeated misbehavior” is a valid measure of insensitivity to discipline, of course, and those who failed this test used to be segregated in “reform schools,” but this practice has fallen out of favor. I am arguing that some fraction of ADHD children probably belong in the “reform school” category.

⁶¹ The traditional self-serving male chauvinist view is of course the opposite.

⁶² David Mamet’s play *Oleanna* presents a vivid picture of this scenario.

⁶³ See Cathy Young, Domestic violations. *Reason*, February, 1998, 23-31.

adult as if he or (more usually) she were a child? This is a topic that requires separate treatment. I mention it here just for completeness.⁶⁴

Identification by "Markers." The third way that individual differences in susceptibility to normal reinforcement contingencies are assessed is via *markers*. Markers can be overt or covert. Overt markers, are usually hard-to-alter visible characteristics correlated with the trait in question. The most obvious examples are sex, age and race. But police "profiles" often use other less invariant features such as dress, grooming, type of car in relation to type of driver, and so on, as ways to maximize the efficiency of "stops". Thus males are stopped more often than females, young males more often than old males and black males more often than white, because men are more likely to commit most crimes than women⁶⁵; young men more likely than old men, etc.

Classification by markers is controversial, perhaps because "race" is one of the markers. What bearing does it have on the concept of responsibility? Just this, that young men are evidently less susceptible to normal deterrents than old men, men less susceptible than women, and so on. The obvious implication is that punishments for young offenders (all other things being equal) should be more severe than for older offenders. This is a conclusion that is probably not acceptable, for reasons that go beyond simple deterrence. Nevertheless, it has guided practice in times past. Fifty years ago in England, for example, so many strokes of the birch were given to young offenders for different classes of crime, but corporal punishment of older men was rare. Notoriously, blacks were in the past subjected to much more severe physical penalties than whites for the same crime, because of a belief that they were less sensitive to punishment. Conversely, the execution of reformed murder Karla Fay Tucker in Texas serves as a reminder that women are even now much less often subjected to extreme penalties than men. The usual explanation offered is "chivalry" by the male establishment. But the practice could equally well be justified on grounds of women's greater sensitivity to negative contingencies, as proved by their lower rates of (most kinds of) crime⁶⁶.

Science has added another dimension to the "marker" question by beginning to discover *covert* features, such as genotype and physiological characteristics, that are correlated with crime. For example, recent work by a variety of neurologists, psychologists and psychiatrists suggests that a history of child abuse coupled with measurable neurological damage strongly predisposes to violent crime.⁶⁷ Psychiatrist Dorothy Lewis and neurologist Jonathan Pincus "believe that the most vicious criminals are, overwhelmingly, people with some combination of abusive childhood, brain injuries, and psychotic symptoms...and that while each of these problems individually has no connection to criminal-

⁶⁴ The entertainment media have certainly picked up on this theme. Two examples: When asked by a TV reporter about his simultaneous affairs with two sisters and their mother, English political charmer, aristocrat and self-confessed rake Alan Clark recently responded "Whatever happened to free will?" The answer seems to be that for women it has been to some degree suspended. And in the popular movie *As Good as it Gets*, misanthrope romance novelist Jack Nicholson, asked how he thinks like a woman, responds "I think of a man. Then I take away reason and accountability." Plus ça change...

⁶⁵ Women tend to lead men in a few crimes, such as poisoning and embezzlement; otherwise, men are well ahead, particularly in violent crime.

⁶⁶ The bald fact of lower crime rates by women leaves open the question of how this comes about. The Darwinian metaphor offers two possibilities that are not mutually exclusive: women are less like to initiate crimes (i.e., show less behavioral variation of this type); or they are more responsive to punishment. The two might be distinguished by comparing the recidivism rate for men and women comparably punished for comparable crimes. (Of course, women might first-offend less often than men because they are more readily deterred by the *prospect* of punishment—i.e., may be show more of the social aspect of responsibility.)

⁶⁷ For a survey see Malcolm Gladwell: Crime and Science: Damaged. *The New Yorker*, Feb. 24, 1997, 132-147.

ity...somehow these factors together create such terrifying synergy as to impede these individuals' ability to play by the rules of society."⁶⁸

Example: an apparently normal individual was offered an orange juice by the nurse in his doctor's office and at once reflexively flung it back at her, damaging her eye, and then when asked why responded "The orange juice was warm" adding "I don't know what got into me." Such a person clearly lacks normal self control, and indeed this particular individual had suffered severe brain damage in an accident; before the accident his behavior had been normal.⁶⁹ On the other hand, another case cited by Gladwell, is serial hate criminal Joseph Paul Franklin. Franklin wounded black civil-rights lawyer Vernon Jordan and shot and paralyzed pornographer Larry Flynt, but also carried out at least one less-famous murder at a synagogue in St. Louis. His preparations for the latter crime were meticulous. He bought the rifle through a for-sale ad (so the purchase could not be traced); he rode a bicycle from his parked car to the synagogue; he filed off the gun's serial number, and so on. His efforts to avoid detection prove that he knew what he was doing was illegal if not wrong – and they also show his fear of judicial punishment. Yet Dorothy Lewis diagnosed him as a "paranoid schizophrenic," basing her diagnosis largely on a history of severe child abuse, and despite his lucidity in interviews and his own view of the insanity defense as "hogwash."

These two cases seem to me very different. In the first case, there is a demonstrable accident, correlated with a clear change in the behavior of the same individual. Moreover, even after the accident, the individual might have been medicated or even trained to restrain himself. Nevertheless, responsibility is clearly diminished. Science might perhaps say how and when such a person is dangerous. Deterrence is clearly not an issue, once the individual is known to be abnormal, normal individuals no longer fall with his deterrence group. Given a known level of danger, what should be done to protect society from such a person is again more a political than a scientific matter.

Franklin, on the other hand, gave every evidence of being sensitive to the contingencies of the law. If punishment had been certain, he would probably not have committed his crimes. He is clearly "abnormal," just because most people don't engage in murder as a hobby. But most people don't keep snakes as pets, which is not to say that herpetologists should be treated differently by the law than everybody else. There are several policy possibilities. *If* we believe that Franklin and others like him can be conclusively identified before they commit their crimes (which is by no means certain), we can institute screening programs to convert these covert markers into overt ones. The best we can hope for, in all probability, is that our tests are about as accurate as the overt markers of age, gender and sex. But then, what do we do? Lock these folk up before they can kill? Does anyone seriously propose that we lock up young, black men, just as a precaution? Perhaps science will bail us out by finding some kind of medication for some of these conditions, although this raises the specter of a Brave-New-World manipulated by drugs⁷⁰. But medications usually have side-effects, and the side-effects are on the individual whereas the benefits of the treatment are conferred on others. Who will make potential criminals take their medicine? Who is responsible if they do not? (The medication problem is already well-known with many homeless people.)

⁶⁸ Gladwell, op. cit., pp. 134-5.

⁶⁹ Gladwell, op. cit., p. 138.

⁷⁰ C. S. Lewis in a prophetic essay, The humanitarian theory of punishment (in *God in the Dock - Essays in Theology and Ethics*. Walter Hooper [ed.], published in 1970, reprinted in 1995. Originally published in: *20th Century: An Australian Quarterly Review*, Vol. III, No. 3, 1948) points out the dangers of medicalizing crime – and argues for a retributive theory of punishment.

The other possibility is that we cannot accurately identify potential criminals in advance. After all, many victims of child abuse and many with damaged brains lead blameless lives. How then should we treat those individuals who do commit horrible crimes? Here the deterrence argument seems to me the overriding one: “‘There was a man,’ remarked Captain Eliot, ‘who was sentenced to death for stealing a horse from a common. He said to the judge that he thought it hard to be hanged for stealing a horse from a common; and the judge answered, ‘You are not to be hanged for stealing a horse from a common, but that others may not steal horses from commons.’”⁷¹

Science and Law

James Q. Wilson, one of our most astute observers on the relations between law and social science, recently took up this issue in his influential little book *Moral Judgment*. First, he describes a scientific criticism of the M’Naghten rule, that it assumes people are either insane or not: “Mental health is not a bipolar state – crazy or sane – but a continuum, a matter of degree.... That criticism is good science but bad law. It is good science because it points out that all behavior is caused; it is bad law because it implies that to the degree a person’s behavior is caused, he is not responsible for it.” And later “The *Durham* decision was the effort of one court to do the impossible—make the law mirror science.”⁷²

The forces that Wilson is battling are real. For example, he goes on to cite Judge David Bazelon, a good friend to mainstream psychology,⁷³ who “In the *Durham* case... said that a person was not responsible for his or her actions if they were the product of a mental disease or defect.”⁷⁴ This change in the M’Naghten rule turned out to be (Surprise! Surprise!) “unworkable” in Wilson’s words. But “Psychiatrists liked the decision; now there was virtually no criminal trial in which their testimony would not be relevant.” Wilson concludes that this will not do, but offers no solution beyond the injunction “that we let neither science nor compassion decide legal precepts.”⁷⁵

Wilson’s comments set up an opposition between science and law that seems to me utterly unjustified. The science to which Wilson refers is not so much bad science, as poor philosophy. The first part of this paper set out the reasons why the concept of responsibility *requires* a certain kind of determinism, namely the capacity to adapt in a predictable way to reinforcement contingencies. The latter part argues that reduced susceptibility to conventional contingencies – one aspect of the fact that mental health is a continuum – can as well lead to increased severity of punishment as to mitigation. Wilson writes: “One may be so depressed, enraged, or panicky as to do things one would not ordinarily do.”⁷⁶ What does science say? Science indicates only that *either* severer penalties may be needed to deter people in one of these conditions, *or* that they should be regarded as mitigating factors. Which is chosen depends not on science but on the values of the culture. For example, suppose a man is enraged because he catches his wife in bed with another man, and shoots the man (or the wife). In the culture of France this used to be thought such an extreme provocation, such a violation of the husband’s rights, that homicide is both understandable and excusable. The husband is not responsible in such a situation;

⁷¹ Patrick O’Brien, *The Mauritius Command*, New York: W. W. Norton, 1977. (P. 210)

⁷² Wilson, J. Q. *Moral judgment: Does the abuse excuse threaten our legal system*. New York: Basic Books. Pp. 36-7. In the *Durham* case of 1954, Monte Durham had received a medical discharge from the Navy for psychiatric reasons. After his discharge, he had been in and out of prisons and mental hospitals for fraud, robbery and other crimes. By current psychiatric standards, Durham, was a psychopath. This case broadened the legal definition of insanity beyond the M’Naghten rule and gave a great boost to testimony from psychiatrists and psychologists.

⁷³ Bazelon, D. L. (1982) Veils, values, and social responsibility. *American Psychologist*, 37, 115-121.

⁷⁴ Wilson, J. Q., *ibid.*, p. 37.

⁷⁵ *Ibid.*, p. 112.

⁷⁶ *Ibid.*, p. 36.

homicide in such cases was treated much like self-defense elsewhere. Verdict: innocent. But in most Anglo-Saxon countries a verdict of at least “manslaughter” would be rendered. The provocation would be considered only partial mitigation.

In modern American culture certain categories of assault, such as wife-battering, are regarded as especially heinous, partially, perhaps, because men are so easily provoked by their wives – the evidence being testimony from women’s groups and a growing number of spousal-abuse cases. The result has been that in many state the penalties are now more severe than for other kinds of assaults: the “up the punishment” option. So-called “hate crimes” are another example where punishment for the same act (e.g., assault) is especially severe if the motive is thought to be racial, ethnic or religious prejudice. There are several reasons for this policy, but one of them is surely the prevailing view that race hatred is an emotion that is especially easily aroused⁷⁷, and so must be met with exceptional deterrence.

What has science to say about these disparities? Why is a predisposition to act taken in one case as mitigation (the self-defense model) and another as reason for greater punishment (the hate-crime model)? The degree of deterministic causation may be the same in both cases. The differences between the legal penalties reflect not scientific but cultural and moral differences. There is not, cannot be, opposition between good science and good law.

CONCLUSION

Responsibility is a behavioristic concept, because responsibilities are discharged by action, not thought. The idea that if behavior is causally determined a man cannot be held responsible for his actions is false. Responsibility depends on a degree of determinism, namely predictable response to contingencies of reward and punishment (*reinforcement contingencies*).

Responsibility, behavioristically defined, has two components: sensitivity to contingencies of reinforcement; and the ability to learn from instruction and example (deterrence). Normal adults possess both components (although there are individual differences). Children, animals and the mentally ill lack one or both.

Contingencies of reinforcement are not determinative. Reinforcement and punishment have direct effects (“misbehavior”) as well as selective effects. Hence, they don’t always work as intended. Consequently, attempts to control crime solely by increasing or decreasing levels of reward and punishment may not work as planned.

The problem of “diminished responsibility” boils down to how we should deal with individual differences in sensitivity to contingencies. Individual differences are either intrinsic, or traceable to external causes such as drugs or upbringing (or both). They can be identified via behavior (repeat offending), behavioral test or overt and covert markers. But in most cases, moral and political⁷⁸ considerations determine whether diminished responsibility is treated as mitigation (e.g., “mental defect”) or leads to increased penalties (e.g., the repeat offender).

Contrary to some authorities, I argue that there can be no conflict between science and law. The role of science is to understand what *is* the case and what is possible. The role of law is to decide what *should be* the case and what should be done, within the limits set by science, politics and morality.

⁷⁷ A wrong view, apparently, as hate crimes are neither common nor increasing in frequency (see James B. Jacobs & Kimberly Potter *Hate crimes: criminal law and identity politics*. Oxford University Press, 1998).

⁷⁸ I don’t add “legal,” because I assume that law is a product of science, politics and morality not an independent entity. But this view may not be general.