

The Model of Rules I

I. EMBARRASSING QUESTIONS

Lawyers lean heavily on the connected concepts of legal right and legal obligation. We say that someone has a legal right or duty, and we take that statement as a sound basis for making claims and demands, and for criticizing the acts of public officials. But our understanding of these concepts is remarkably fragile, and we fall into trouble when we try to say what legal rights and obligations are. We say glibly that whether someone has a legal obligation is determined by applying 'the law' to the particular facts of his case, but this is not a helpful answer, because we have the same difficulties with the concept of law.

We are used to summing up our troubles in the classic questions of jurisprudence: What is 'the law'? When two sides disagree, as often happens, about a proposition 'of law', what are they disagreeing about, and how shall we decide which side is right? Why do we call what 'the law' says a matter of legal 'obligation'? Is 'obligation' here just a term of art, meaning only what the law says? Or does legal obligation have something to do with moral obligation? Can we say that we have, in principle at least, the same reasons for meeting our legal obligations that we have for meeting our moral obligations?

These are not puzzles for the cupboard, to be taken down on rainy days for fun. They are sources of continuing embarrassment, and they nag at our attention. They embarrass us in dealing with particular problems that we must solve, one way or another. Suppose a novel right-of-privacy case comes to court, and there is no statute or precedent claimed by the plaintiff. What role in the court's decision should be played by the fact that most people in the community think that private individuals are 'morally' entitled to that particular privacy? Supposing the Supreme Court orders some prisoner freed because the police used procedures that the Court now says are constitutionally forbidden, although the Court's earlier decisions upheld these procedures. Must the Court, to be consistent, free all other prisoners previously convicted through these same procedures?¹ Conceptual puzzles about 'the law' and 'legal obligation'

¹ See *Linkletter v. Walker*, 381 U.S. 618 (1965).

become acute when a court is confronted with a problem like this.

These eruptions signal a chronic disease. Day in and day out we send people to jail, or take money away from them, or make them do things they do not want to do, under coercion of force, and we justify all of this by speaking of such persons as having broken the law or having failed to meet their legal obligations, or having interfered with other people's legal rights. Even in clear cases (a bank robber or a wilful breach of contract), when we are confident that someone had a legal obligation and broke it, we are not able to give a satisfactory account of what that means, or why that entitles the state to punish or coerce him. We may feel confident that what we are doing is proper, but until we can identify the principles we are following we cannot be sure that they are sufficient, or whether we are applying them consistently. In less clear cases, when the issue of whether an obligation has been broken is for some reason controversial, the pitch of these nagging questions rises, and our responsibility to find answers deepens.

Certain lawyers (we may call them 'nominalists') urge that we solve these problems by ignoring them. In their view the concepts of 'legal obligation' and 'the law' are myths, invented and sustained by lawyers for a dismal mix of conscious and subconscious motives. The puzzles we find in these concepts are merely symptoms that they are myths. They are unsolvable because unreal, and our concern with them is just one feature of our enslavement. We would do better to flush away the puzzles and the concepts altogether, and pursue our important social objectives without this excess baggage.

This is a tempting suggestion, but it has fatal drawbacks. Before we can decide that our concepts of law and of legal obligation are myths, we must decide what they are. We must be able to state, at least roughly, what it is we all believe that is wrong. But the nerve of our problem is that we have great difficulty in doing just that. Indeed, when we ask what law is and what legal obligations are, we are asking for a theory of how we use those concepts and of the conceptual commitments our use entails. We cannot conclude, before we have such a general theory, that our practices are stupid or superstitious.

Of course, the nominalists think they know how the rest of us use these concepts. They think that when we speak of 'the law' we mean a set of timeless rules stocked in some conceptual warehouse awaiting discovery by judges, and that when we speak of legal obligation we mean the invisible chains these mysterious rules somehow drape around us. The theory that there are such rules and chains they call 'mechanical jurisprudence', and they are right in ridiculing its practitioners. Their difficulty, however, lies in finding practitioners to ridicule. So far they have had little luck in caging and exhibiting mechanical jurists.

(all specimens captured – even Blackstone and Joseph Beale – have had to be released after careful reading of their texts).

In any event, it is clear that most lawyers have nothing like this in mind when they speak of the law and of legal obligation. A superficial examination of our practices is enough to show this for we speak of laws changing and evolving, and of legal obligation sometimes being problematical. In these and other ways we show that we are not addicted to mechanical jurisprudence.

Nevertheless, we do use the concepts of law and legal obligation, and we do suppose that society's warrant to punish and coerce is written in that currency. It may be that when the details of this practice are laid bare, the concepts we do use will be shown to be as silly and as thick with illusion as those the nominalists invented. If so, then we shall have to find other ways to describe what we do, and either provide other justifications or change our practices. But until we have discovered this and made these adjustments, we cannot accept the nominalists' premature invitation to turn our backs on the problems our present concepts provide.

Of course the suggestion that we stop talking about 'the law' and 'legal obligation' is mostly bluff. These concepts are too deeply cemented into the structure of our political practices – they cannot be given up like cigarettes or hats. Some of the nominalists have half-admitted this and said that the myths they condemn should be thought of as Platonic myths and retained to seduce the masses into order. This is perhaps not so cynical a suggestion as it seems; perhaps it is a covert hedging of a dubious bet.

If we boil away the bluff, the nominalist attack reduces to an attack on mechanical jurisprudence. Through the lines of the attack, and in spite of the heroic calls for the death of law, the nominalists themselves have offered an analysis of how the terms 'law' and 'legal obligation' should be used which is not very different from that of more classical philosophers. Nominalists present their analysis as a model of how legal institutions (particularly courts) 'really operate'. But their model differs mainly in emphasis from the theory first made popular by the nineteenth century philosopher John Austin, and now accepted in one form or another by most working and academic lawyers who hold views on jurisprudence. I shall call this theory, with some historical looseness, 'legal positivism'. I want to examine the soundness of legal positivism, particularly in the powerful form that Professor H. L. A. Hart has given to it. I choose to focus on his position, not only because of its clarity and elegance, but because here, as almost everywhere else in legal philosophy, constructive thought must start with a consideration of his views.

2. POSITIVISM

Positivism has a few central and organizing propositions as its skeleton, and though not every philosopher who is called a positivist would subscribe to these in the way I present them, they do define the general position I want to examine. These key tenets may be stated as follows:

(a) The law of a community is a set of special rules used by the community directly or indirectly for the purpose of determining which behavior will be punished or coerced by the public power. These special rules can be identified and distinguished by specific criteria, by tests having to do not with their content but with their *pedigree* or the manner in which they were adopted or developed. These tests of pedigree can be used to distinguish valid legal rules from spurious legal rules (rules which lawyers and litigants wrongly argue are rules of law) and also from other sorts of social rules (generally lumped together as 'moral rules') that the community follows but does not enforce through public power.

(b) The set of these valid legal rules is exhaustive of 'the law', so that if someone's case is not clearly covered by such a rule (because there is none that seems appropriate, or those that seem appropriate are vague, or for some other reason) then that case cannot be decided by 'applying the law.' It must be decided by some official, like a judge, 'exercising his discretion,' which means reaching beyond the law for some other sort of standard to guide him in manufacturing a fresh legal rule or supplementing an old one.

(c) To say that someone has a 'legal obligation' is to say that his case falls under a valid legal rule that requires him to do or to forbear from doing something. (To say he has a legal right, or has a legal power of some sort, or a legal privilege or immunity, is to assert, in a shorthand way, that others have actual or hypothetical legal obligations to act or not to act in certain ways touching him.) In the absence of such a valid legal rule there is no legal obligation; it follows that when the judge decides an issue by exercising his discretion, he is not enforcing a legal right as to that issue.

This is only the skeleton of positivism. The flesh is arranged differently by different positivists, and some even tinker with the bones. Different versions differ chiefly in their description of the fundamental test of pedigree a rule must meet to count as a rule of law.

Austin, for example, framed his version of the fundamental test as a series of interlocking definitions and distinctions.¹ He defined having an obligation as lying under a rule, a rule as a general command, and a command as an expression of desire that others behave in a particular way, backed by the power and will to enforce that expression in the event of disobedience. He distinguished classes of rules (legal, moral or religious) according to which person or group is the author of the general command the rule represents. In each political community, he thought, one will find a sovereign – a person or a determinate group whom the rest obey habitually, but who is not in the habit of obeying anyone else. The legal rules of a community are the general commands its sovereign has deployed. Austin's definition of legal obligation followed from this definition of law. One has a legal obligation, he thought, if one is among the addressees of some general order of the sovereign, and is in danger of suffering a sanction unless he obeys that order.

Of course, the sovereign cannot provide for all contingencies through any scheme of orders, and some of his orders will inevitably be vague or have fuzzy edges. Therefore (according to Austin) the sovereign grants those who enforce the law (judges) discretion to make fresh orders when novel or troublesome cases are presented. The judges then make new rules or adapt old rules, and the sovereign either overturns their creations or tacitly confirms them by failing to do so.

Austin's model is quite beautiful in its simplicity. It asserts the first tenet of positivism, that the law is a set of rules specially selected to govern public order, and offers a simple factual test – what has the sovereign commanded? – as the sole criterion for identifying those special rules. In time, however, those who studied and tried to apply Austin's model found it too simple. Many objections were raised, among which were two that seemed fundamental. First, Austin's key assumption that in each community a determinate group or institution can be found, which is in ultimate control of all other groups, seemed not to hold in a complex society. Political control in a modern nation is pluralistic and shifting, a matter of more or less, of compromise and cooperation and alliance, so that it is often impossible to say that any person or group has that dramatic control necessary to qualify as an Austinian sovereign. One wants to say, in the United States for example, that the 'people' are sovereign. But this means almost nothing, and in itself provides no test for determining what the 'people' have commanded, or distinguishing their legal from their social or moral commands.

Second, critics began to realize that Austin's analysis fails entirely to account for, even to recognize, certain striking facts about the attitudes

we take toward 'the law.' We make an important distinction between law and even the general orders of a gangster. We feel that the law's strictures – and its sanctions – are different in that they are obligatory in a way that the outlaw's commands are not. Austin's analysis has no place for any such distinction, because it defines an obligation as subjection to the threat of force, and so founds the authority of law entirely on the sovereign's ability and will to harm those who disobey. Perhaps the distinction we make is illusory – perhaps our feelings of some special authority attaching to the law is based on religious hangover or another sort of mass self-deception. But Austin does not demonstrate this, and we are entitled to insist that an analysis of our concept of law either acknowledge and explain our attitudes, or show why they are mistaken.

H. L. A. Hart's version of positivism is more complex than Austin's, in two ways. First, he recognizes, as Austin did not, that rules are of different logical kinds. (Hart distinguishes two kinds, which he calls 'primary' and 'secondary' rules). Second, he rejects Austin's theory that a rule is a kind of command, and substitutes a more elaborate general analysis of what rules are. We must pause over each of these points, and then note how they merge in Hart's concept of law.

Hart's distinction between primary and secondary rules is of great importance.¹ Primary rules are those that grant rights or impose obligations upon members of the community. The rules of the criminal law that forbid us to rob, murder or drive too fast are good examples of primary rules. Secondary rules are those that stipulate how, and by whom, such primary rules may be formed, recognized, modified or extinguished. The rules that stipulate how Congress is composed, and how it enacts legislation, are examples of secondary rules. Rules about forming contracts and executing wills are also secondary rules because they stipulate how very particular rules governing particular legal obligations (*i.e.*, the terms of a contract or the provisions of a will) come into existence and are changed.

His general analysis of rules is also of great importance.² Austin had said that every rule is a general command, and that a person is obligated under a rule if he is liable to be hurt should he disobey it. Hart points out that this obliterates the distinction between being *obliged* to do something and being *obligated* to do it. If one is bound by a rule he is obligated, not merely obliged, to do what it provides, and therefore being bound by a rule must be different from being subject to an injury if one disobeys an order. A rule differs from an order, among other ways, by being *normative*, by setting a standard of behavior that has a call on

¹ See H. L. A. Hart, *The Concept of Law*, 89-96 (1961).

² *Id.* at 79-88.

¹ J. Austin, *The Province of Jurisprudence Determined* (1832).

its subject beyond the threat that may enforce it. A rule can never be binding just because some person with physical power wants it to be so. He must have *authority* to issue the rule or it is no rule, and such authority can only come from another rule which is already binding on those to whom he speaks. That is the difference between a valid law and the orders of a gunman.

So Hart offers a general theory of rules that does not make their authority depend upon the physical power of their authors. If we examine the way different rules come into being, he tells us, and attend to the distinction between primary and secondary rules, we see that there are two possible sources of a rule's authority:¹

(a) A rule may become binding upon a group of people because that group through its practices *accepts* the rule as a standard for its conduct. It is not enough that the group simply conforms to a pattern of behavior: even though most Englishmen may go to the movies on Saturday evening, they have not accepted a rule requiring that they do so. A practice constitutes the acceptance of a rule only when those who follow the practice regard the rule as binding, and recognize the rule as a reason or justification for their own behavior and as a reason for criticizing the behavior of others who do not obey it.

(b) A rule may also become binding in quite a different way, namely by being enacted in conformity with some *secondary* rule that stipulates that rules so enacted shall be binding. If the constitution of a club stipulates, for example, that by-laws may be adopted by a majority of the members, then particular by-laws so voted are binding upon all the members, not because of any practice of acceptance of these particular by-laws, but because the constitution says so. We use the concept of *validity* in this connection: rules binding because they have been created in a manner stipulated by some secondary rule are called 'valid' rules.

Thus we can record Hart's fundamental distinction this way: a rule may be binding (a) because it is accepted or (b) because it is valid.

Hart's concept of law is a construction of these various distinctions.² Primitive communities have only primary rules, and these are binding entirely because of practices of acceptance. Such communities cannot be said to have 'law,' because there is no way to distinguish a set of legal rules from amongst other social rules, as the first tenet of positivism

¹ *Id.* at 97-107.

² *Id. passim*, particularly ch. 6.

requires. But when a particular community has developed a fundamental secondary rule that stipulates how legal rules are to be identified, the idea of a distinct set of legal rules, and thus of law, is born.

Hart calls such a fundamental secondary rule a 'rule of recognition'. The rule of recognition of a given community may be relatively simple ('What the king enacts is law') or it may be very complex (the United States Constitution, with all its difficulties of interpretation, may be considered a single rule of recognition). The demonstration that a particular rule is valid may therefore require tracing a complicated chain of validity back from that particular rule ultimately to the fundamental rule. Thus a parking ordinance of the city of New Haven is valid because it is adopted by a city council, pursuant to the procedures and within the competence specified by the municipal law adopted by the state of Connecticut, in conformity with the procedures and within the competence specified by the constitution of the state of Connecticut, which was in turn adopted consistently with the requirements of the United States Constitution.

Of course, a rule of recognition cannot itself be valid, because by hypothesis it is ultimate, and so cannot meet tests stipulated by a more fundamental rule. The rule of recognition is the sole rule in a legal system whose binding force depends upon its acceptance. If we wish to know what rule of recognition a particular community has adopted or follows, we must observe how its citizens, and particularly its officials, behave. We must observe what ultimate arguments they accept as showing the validity of a particular rule, and what ultimate arguments they use to criticize other officials or institutions. We can apply no mechanical test, but there is no danger of our confusing the rule of recognition of a community with its rules of morality. The rule of recognition is identified by the fact that its province is the operation of the governmental apparatus of legislatures, courts, agencies, policemen, and the rest.

In this way Hart rescues the fundamentals of positivism from Austin's mistakes. Hart agrees with Austin that valid rules of law may be created through the acts of officials and public institutions. But Austin thought that the authority of these institutions lay only in their monopoly of power. Hart finds their authority in the background of constitutional standards against which they act, constitutional standards that have been accepted, in the form of a fundamental rule of recognition, by the community which they govern. This background legitimates the decisions of government and gives them the cast and call of obligation that the naked commands of Austin's sovereign lacked. Hart's theory differs from Austin's also, in recognizing that different communities use different ultimate tests of law, and that some allow other means of creating law than the deliberate act of a legislative institution. Hart mentions 'long

customary practice' and 'the relation [of a rule] to judicial decisions' as other criteria that are often used, though generally along with and subordinate to the test of legislation.

So Hart's version of positivism is more complex than Austin's, and his test for valid rules of law is more sophisticated. In one respect, however, the two models are very similar. Hart, like Austin, recognizes that legal rules have furry edges (he speaks of them as having 'open texture') and, again like Austin, he accounts for troublesome cases by saying that judges have and exercise discretion to decide these cases by fresh legislation.¹ (I shall later try to show why one who thinks of law as a special set of rules is almost inevitably drawn to account for difficult cases in terms of someone's exercise of discretion.)

3. RULES, PRINCIPLES, AND POLICIES

I want to make a general attack on positivism, and I shall use H. L. A. Hart's version as a target, when a particular target is needed. My strategy will be organized around the fact that when lawyers reason or dispute about legal rights and obligations, particularly in those hard cases when our problems with these concepts seem most acute, they make use of standards that do not function as rules, but operate differently as principles, policies, and other sorts of standards. Positivism, I shall argue, is a model of and for a system of rules, and its central notion of a single fundamental test for law forces us to miss the important roles of these standards that are not rules.

I just spoke of 'principles, policies, and other sorts of standards'. Most often I shall use the term 'principle' generically, to refer to the whole set of these standards other than rules; occasionally, however, I shall be more precise, and distinguish between principles and policies. Although nothing in the present argument will turn on the distinction, I should state how I draw it. I call a 'policy' that kind of standard that sets out a goal to be reached, generally an improvement in some economic, political, or social feature of the community (though some goals are negative, in that they stipulate that some present feature is to be protected from adverse change). I call a 'principle' a standard that is to be observed, not because it will advance or secure an economic, political, or social situation deemed desirable, but because it is a requirement of justice or fairness or some other dimension of morality. Thus the standard that automobile accidents are to be decreased is a policy, and the standard that no man may profit by his own wrong a principle. The distinction can be collapsed by construing a principle as stating a social goal (*i.e.*, the goal of a

society in which no man profits by his own wrong), or by construing a policy as stating a principle (*i.e.*, the principle that the goal the policy embraces is a worthy one) or by adopting the utilitarian thesis that principles of justice are disguised statements of goals (securing the greatest happiness of the greatest number). In some contexts the distinction has uses which are lost if it is thus collapsed.¹

My immediate purpose, however, is to distinguish principles in the generic sense from rules, and I shall start by collecting some examples of the former. The examples I offer are chosen haphazardly; almost any case in a law school casebook would provide examples that would serve as well. In 1889 a New York court, in the famous case of *Riggs v. Palmer*,² had to decide whether an heir named in the will of his grandfather could inherit under that will, even though he had murdered his grandfather to do so. The court began its reasoning with this admission: 'It is quite true that statutes regulating the making, proof and effect of wills, and the devolution of property, if literally construed, and if their force and effect can in no way and under no circumstances be controlled or modified, give this property to the murderer.'³ But the court continued to note that 'all laws as well as all contracts may be controlled in their operation and effect by general, fundamental maxims of the common law. No one shall be permitted to profit by his own fraud, or to take advantage of his own wrong, or to found any claim upon his own iniquity, or to acquire property by his own crime.'⁴ The murderer did not receive his inheritance.

In 1960, a New Jersey court was faced, in *Henningsen v. Bloomfield Motors, Inc.*⁵ with the important question of whether (or how much) an automobile manufacturer may limit his liability in case the automobile is defective. Henningsen had bought a car, and signed a contract which said that the manufacturer's liability for defects was limited to 'making good' defective parts – 'this warranty being expressly in lieu of all other warranties, obligations or liabilities.' Henningsen argued that, at least in the circumstances of his case, the manufacturer ought not to be protected by this limitation, and ought to be liable for the medical and other expenses of persons injured in a crash. He was not able to point to any statute, or to any established rule of law, that prevented the manufacturer from standing on the contract. The court nevertheless agreed with Henningsen. At various points in the court's argument the

¹ See Chapter 4. See also Dworkin, 'Wasserstrom: The Judicial Decision', 75 *Ethics* 47 (1964), reprinted as 'Does Law Have a Function?', 74 *Yale Law Journal* 640 (1965).

² 115 N.Y. 506, 22 N.E. 188 (1889).

³ *Id.* at 509, 22 N.E. at 189.

⁴ *Id.* at 511, 22 N.E. at 190.

⁵ 32 N.J. 358, 161 A.2d 69 (1960).

following appeals to standards are made: (a) '[W]e must keep in mind the general principle that, in the absence of fraud, one who does not choose to read a contract before signing it cannot later relieve himself of its burdens.'¹ (b) 'In applying that principle, the basic tenet of freedom of competent parties to contract is a factor of importance.'² (c) 'Freedom of contract is not such an immutable doctrine as to admit of no qualification in the area in which we are concerned.'³ (d) 'In a society such as ours, where the automobile is a common and necessary adjunct of daily life, and where its use is so fraught with danger to the driver, passengers and the public, the manufacturer is under a special obligation in connection with the construction, promotion and sale of his cars. Consequently, the courts must examine purchase agreements closely to see if consumer and public interests are treated fairly.'⁴ (e) '“[I]s there any principle which is more familiar or more firmly embedded in the history of Anglo-American law than the basic doctrine that the courts will not permit themselves to be used as instruments of inequity and injustice?”'⁵ (f) '“More specifically the courts generally refuse to lend themselves to the enforcement of a ‘bargain’ in which one party has unjustly taken advantage of the economic necessities of other”'⁶

The standards set out in these quotations are not the sort we think of as legal rules. They seem very different from propositions like 'The maximum legal speed on the turnpike is sixty miles an hour' or 'A will is invalid unless signed by three witnesses'. They are different because they are legal principles rather than legal rules.

The difference between legal principles and legal rules is a logical distinction. Both sets of standards point to particular decisions about legal obligation in particular circumstances, but they differ in the character of the direction they give. Rules are applicable in an all-or-nothing fashion. If the facts a rule stipulates are given, then either the rule is valid, in which case the answer it supplies must be accepted, or it is not, in which case it contributes nothing to the decision.

This all-or-nothing is seen most plainly if we look at the way rules operate, not in law, but in some enterprise they dominate – a game, for example. In baseball a rule provides that if the batter has had three strikes, he is out. An official cannot consistently acknowledge that this is an accurate statement of a baseball rule, and decide that a batter who has had three strikes is not out. Of course, a rule may have exceptions

¹ *Id.* at 386, 161 A.2d at 84.

² *Id.*

³ *Id.* at 388, 161 A.2d at 86.

⁴ *Id.* at 387, 161 A.2d at 85.

⁵ *Id.* at 389, 161 A.2d at 86 (quoting Frankfurter, J., in *United States v. Bethlehem Steel*, 315 U.S. 289, 326 [1942]).

⁶ *Id.*

(the batter who has taken three strikes is not out if the catcher drops the third strike). However, an accurate statement of the rule would take this exception into account, and any that did not would be incomplete. If the list of exceptions is very large, it would be too clumsy to repeat them each time the rule is cited; there is, however, no reason in theory why they could not all be added on, and the more that are, the more accurate is the statement of the rule.

If we take baseball rules as a model, we find that rules of law, like the rule that a will is invalid unless signed by three witnesses, fit the model well. If the requirement of three witnesses is a valid legal rule, then it cannot be that a will has been signed by only two witnesses and is valid. The rule might have exceptions, but if it does then it is inaccurate and incomplete to state the rule so simply, without enumerating the exceptions. In theory, at least, the exceptions could all be listed, and the more of them that are, the more complete is the statement of the rule.

But this is not the way the sample principles in the quotations operate. Even those which look most like rules do not set out legal consequences that follow automatically when the conditions provided are met. We say that our law respects the principle that no man may profit from his own wrong, but we do not mean that the law never permits a man to profit from wrongs he commits. In fact, people often profit, perfectly legally, from their legal wrongs. The most notorious case is adverse possession – if I trespass on your land long enough, some day I will gain a right to cross your land whenever I please. There are many less dramatic examples. If a man leaves one job, breaking a contract, to take a much higher paying job, he may have to pay damages to his first employer, but he is usually entitled to keep his new salary. If a man jumps bail and crosses state lines to make a brilliant investment in another state, he may be sent back to jail, but he will keep his profits.

We do not treat these – and countless other counter-instances that can easily be imagined – as showing that the principle about profiting from one's wrongs is not a principle of our legal system, or that it is incomplete and needs qualifying exceptions. We do not treat counter-instances as exceptions (at least not exceptions in the way in which a catcher's dropping the third strike is an exception) because we could not hope to capture these counter-instances simply by a more extended statement of the principle. They are not, even in theory, subject to enumeration, because we would have to include not only these cases (like adverse possession) in which some institution has already provided that profit can be gained through a wrong, but also those numberless imaginary cases in which we know in advance that the principle would not hold. Listing some of these might sharpen our sense of the principle's weight (I shall mention that dimension in a moment), but it would not

make for a more accurate or complete statement of the principle.

A principle like 'No man may profit from his own wrong' does not even purport to set out conditions that make its application necessary. Rather, it states a reason that argues in one direction, but does not necessitate a particular decision. If a man has or is about to receive something, as a direct result of something illegal he did to get it, then that is a reason which the law will take into account in deciding whether he should keep it. There may be other principles or policies arguing in the other direction – a policy of securing title, for example, or a principle limiting punishment to what the legislature has stipulated. If so, our principle may not prevail, but that does not mean that it is not a principle of our legal system, because in the next case, when these contravening considerations are absent or less weighty, the principle may be decisive. All that is meant, when we say that a particular principle is a principle of our law, is that the principle is one which officials must take into account, if it is relevant, as a consideration inclining in one direction or another.

The logical distinction between rules and principles appears more clearly when we consider principles that do not even look like rules. Consider the proposition, set out under '(d)' in the excerpts from the *Henningsen* opinion, that 'the manufacturer is under a special obligation in connection with the construction, promotion and sale of his cars'. This does not even purport to define the specific duties such a special obligation entails, or to tell us what rights automobile consumers acquire as a result. It merely states – and this is an essential link in the *Henningsen* argument – that automobile manufacturers must be held to higher standards than other manufacturers, and are less entitled to rely on the competing principle of freedom of contract. It does not mean that they may never rely on that principle, or that courts may rewrite automobile purchase contracts at will; it means only that if a particular clause seems unfair or burdensome, courts have less reason to enforce the clause than if it were for the purchase of neckties. The 'special obligation' counts in favor, but does not in itself necessitate, a decision refusing to enforce the terms of an automobile purchase contract.

This first difference between rules and principles entails another. Principles have a dimension that rules do not – the dimension of weight or importance. When principles intersect (the policy of protecting automobile consumers intersecting with principles of freedom of contract, for example), one who must resolve the conflict has to take into account the relative weight of each. This cannot be, of course, an exact measurement, and the judgment that a particular principle or policy is more important than another will often be a controversial one. Nevertheless, it is an integral part of the concept of a principle that it has this dimension,

that it makes sense to ask how important or how weighty it is.

Rules do not have this dimension. We can speak of rules as being *functionally* important or unimportant (the baseball rule that three strikes are out is more important than the rule that runners may advance on a balk, because the game would be much more changed with the first rule altered than the second). In this sense, one legal rule may be more important than another because it has a greater or more important role in regulating behavior. But we cannot say that one rule is more important than another within the system of rules, so that when two rules conflict one supersedes the other by virtue of its greater weight.

If two rules conflict, one of them cannot be a valid rule. The decision as to which is valid, and which must be abandoned or recast, must be made by appealing to considerations beyond the rules themselves. A legal system might regulate such conflicts by other rules, which prefer the rule enacted by the higher authority, or the rule enacted later, or the more specific rule, or something of that sort. A legal system may also prefer the rule supported by the more important principles. (Our own legal system uses both of these techniques.)

It is not always clear from the form of a standard whether it is a rule or a principle. 'A will is invalid unless signed by three witnesses' is not very different in form from 'A man may not profit from his own wrong', but one who knows something of American law knows that he must take the first as stating a rule and the second as stating a principle. In many cases the distinction is difficult to make – it may not have been settled how the standard should operate, and this issue may itself be a focus of controversy. The first amendment to the United States Constitution contains the provision that Congress shall not abridge freedom of speech. Is this a rule, so that if a particular law does abridge freedom of speech, it follows that it is unconstitutional? Those who claim that the first amendment is 'an absolute' say that it must be taken in this way, that is, as a rule. Or does it merely state a principle, so that when an abridgement of speech is discovered, it is unconstitutional unless the context presents some other policy or principle which in the circumstances is weighty enough to permit the abridgement? That is the position of those who argue for what is called the 'clear and present danger' test or some other form of 'balancing'.

Sometimes a rule and a principle can play much the same role, and the difference between them is almost a matter of form alone. The first section of the Sherman Act states that every contract in restraint of trade shall be void. The Supreme Court had to make the decision whether this provision should be treated as a rule in its own terms (striking down every contract 'which restrains trade', which almost any contract does) or as a principle, providing a reason for striking down a contract in the

absence of effective contrary policies. The Court construed the provision as a rule, but treated that rule as containing the word 'unreasonable', and as prohibiting only 'unreasonable' restraints of trade.¹ This allowed the provision to function logically as a rule (whenever a court finds that the restraint is 'unreasonable' it is bound to hold the contract invalid) and substantially as a principle (a court must take into account a variety of other principles and policies in determining whether a particular restraint in particular economic circumstances is 'unreasonable').

Words like 'reasonable', 'negligent', 'unjust', and 'significant' often perform just this function. Each of these terms makes the application of the rule which contains it depend to some extent upon principles or policies lying beyond the rule, and in this way makes that rule itself more like a principle. But they do not quite turn the rule into a principle, because even the least confining of these terms restricts the *kind* of other principles and policies on which the rule depends. If we are bound by a rule that says that 'unreasonable' contracts are void, or that grossly 'unfair' contracts will not be enforced, much more judgment is required than if the quoted terms were omitted. But suppose a case in which some consideration of policy or principle suggests that a contract should be enforced even though its restraint is not reasonable, or even though it is grossly unfair. Enforcing these contracts would be forbidden by our rules, and thus permitted only if these rules were abandoned or modified. If we were dealing, however, not with a rule but with a policy against enforcing unreasonable contracts, or a principle that unfair contracts ought not to be enforced, the contracts could be enforced without alteration of the law.

4. PRINCIPLES AND THE CONCEPT OF LAW

Once we identify legal principles as separate sorts of standards, different from legal rules, we are suddenly aware of them all around us. Law teachers teach them, lawbooks cite them, legal historians celebrate them. But they seem most energetically at work, carrying most weight, in difficult lawsuits like *Riggs* and *Henningsen*. In cases like these, principles play an essential part in arguments supporting judgments about particular legal rights and obligations. After the case is decided, we may say that the case stands for a particular rule (e.g., the rule that one who murders is not eligible to take under the will of his victim). But the rule does not exist before the case is decided; the court cites principles as its justification for adopting and applying a new rule. In *Riggs*, the court

cited the principle that no man may profit from his own wrong as a background standard against which to read the statute of wills and in this way justified a new interpretation of that statute. In *Henningsen*, the court cited a variety of intersecting principles and policies as authority for a new rule respecting manufacturers' liability for automobile defects.

An analysis of the concept of legal obligation must therefore account for the important role of principles in reaching particular decisions of law. There are two very different tacks we might take:

(a) We might treat legal principles the way we treat legal rules and say that some principles are binding as law and must be taken into account by judges and lawyers who make decisions of legal obligation. If we took this tack, we should say that in the United States, at least, the 'law' includes principles as well as rules.

(b) We might, on the other hand, deny that principles can be binding the way some rules are. We would say, instead, that in cases like *Riggs* or *Henningsen* the judge reaches beyond the rules that he is bound to apply (reaches, that is, beyond the 'law') for extra-legal principles he is free to follow if he wishes.

One might think that there is not much difference between these two lines of attack, that it is only a verbal question of how one wants to use the word 'law'. But that is a mistake, because the choice between these two accounts has the greatest consequences for an analysis of legal obligation. It is a choice between two *concepts* of a legal principle, a choice we can clarify by comparing it to a choice we might make between two concepts of a legal rule. We sometimes say of someone that he 'makes it a rule' to do something, when we mean that he has chosen to follow a certain practice. We might say that someone has made it a rule, for example, to run a mile before breakfast because he wants to be healthy and believes in a regimen. We do not mean, when we say this, that he is *bound* by the rule that he must run a mile before breakfast, or even that he regards it as binding upon him. Accepting a rule as binding is something different from making it a rule to do something. If we use Hart's example again, there is a difference between saying that Englishmen make it a rule to see a movie once a week, and saying that the English have a rule that one must see a movie once a week. The second implies that if an Englishman does not follow the rule, he is subject to criticism or censure, but the first does not. The first does not exclude the possibility of a *sort* of criticism – we can say that one who does not see movies is neglecting his education – but we do not suggest

¹ *Standard Oil v. United States*, 221 U.S. 1, 60 (1911); *United States v. American Tobacco Co.*, 221 U.S. 106, 180 (1911).

that he is doing something wrong *just* in not following the rule.¹

If we think of the judges of a community as a group, we could describe the rules of law they follow in these two different ways. We could say, for instance, that in a certain state the judges make it a rule not to enforce wills unless there are three witnesses. This would not imply that the rare judge who enforces such a will is doing anything wrong just for that reason. On the other hand we can say that in that state a rule of law requires judges not to enforce such wills; this does imply that a judge who enforces them is doing something wrong. Hart, Austin and other positivists, of course, would insist on this latter account of legal rules; they would not at all be satisfied with the 'make it a rule' account. It is not a verbal question of which account is right. It is a question of which describes the social situation more accurately. Other important issues turn on which description we accept. If judges simply 'make it a rule' not to enforce certain contracts, for example, then we cannot say, before the decision, that anyone is 'entitled' to that result, and that proposition cannot enter into any justification we might offer for the decision.

The two lines of attack on principles parallel these two accounts of rules. The first tack treats principles as binding upon judges, so that they are wrong not to apply the principles when they are pertinent. The second tack treats principles as summaries of what most judges 'make it a principle' to do when forced to go beyond the standards that bind them. The choice between these approaches will affect, perhaps even determine, the answer we can give to the question whether the judge in a hard case like *Riggs* or *Henningsen* is attempting to enforce pre-existing legal rights and obligations. If we take the first tack, we are still free to argue that because such judges are applying binding legal standards they are enforcing legal rights and obligations. But if we take the second, we are out of court on that issue, and we must acknowledge that the murderer's family in *Riggs* and the manufacturer in *Henningsen* were deprived of their property by an act of judicial discretion applied *ex post facto*. This may not shock many readers – the notion of judicial discretion has percolated through the legal community – but it does illustrate one of the most nettlesome of the puzzles that drive philosophers to worry about legal obligation. If taking property away in cases like these cannot be justified by appealing to an established obligation, another justification must be found, and nothing satisfactory has yet been supplied.

In my skeleton diagram of positivism, previously set out, I listed the doctrine of judicial discretion as the second tenet. Positivists hold that

¹ The distinction is in substance the same as that made by Rawls, 'Two Concepts of Rules', 64 *Philosophical Review* 3 (1955).

when a case is not covered by a clear rule, a judge must exercise his discretion to decide that case by what amounts to a fresh piece of legislation. There may be an important connection between this doctrine and the question of which of the two approaches to legal principles we must take. We shall therefore want to ask whether the doctrine is correct, and whether it implies the second approach, as it seems on its face to do. En route to these issues, however, we shall have to polish our understanding of the concept of discretion. I shall try to show how certain confusions about that concept and in particular a failure to discriminate different senses in which it is used, account for the popularity of the doctrine of discretion. I shall argue that in the sense in which the doctrine does have a bearing on our treatment of principles, it is entirely unsupported by the arguments the positivists use to defend it.

5. DISCRETION

The concept of discretion was lifted by the positivists from ordinary language, and to understand it we must put it back in *habitat* for a moment. What does it mean, in ordinary life, to say that someone 'has discretion?' The first thing to notice is that the concept is out of place in all but very special contexts. For example, you would not say that I either do or do not have discretion to choose a house for my family. It is not true that I have 'no discretion' in making that choice, and yet it would be almost equally misleading to say that I do have discretion. The concept of discretion is at home in only one sort of context; when someone is in general charged with making decisions subject to standards set by a particular authority. It makes sense to speak of the discretion of a sergeant who is subject to orders of superiors, or the discretion of a sports official or contest judge who is governed by a rule book or the terms of the contest. Discretion, like the hole in a doughnut, does not exist except as an area left open by a surrounding belt of restriction. It is therefore a relative concept. It always makes sense to ask, 'Discretion under which standards?' or 'Discretion as to which authority?' Generally the context will make the answer to this plain, but in some cases the official may have discretion from one stand-point though not from another.

Like almost all terms, the precise meaning of 'discretion' is affected by features of the context. The term is always colored by the background of understood information against which it is used. Although the shadings are many, it will be helpful for us to recognize some gross distinctions.

Sometimes we use 'discretion' in a weak sense, simply to say that for some reason the standards an official must apply cannot be applied mechanically but demand the use of judgment. We use this weak sense

when the context does not already make that clear, when the background our audience assumes does not contain that piece of information. Thus we might say, 'The sergeant's orders left him a great deal of discretion', to those who do not know what the sergeant's orders were or who do not know something that made those orders vague or hard to carry out. It would make perfect sense to add, by way of amplification, that the lieutenant had ordered the sergeant to take his five most experienced men on patrol but that it was hard to determine which were the most experienced.

Sometimes we use the term in a different weak sense, to say only that some official has final authority to make a decision and cannot be reviewed and reversed by any other official. We speak this way when the official is part of a hierarchy of officials structured so that some have higher authority but in which the patterns of authority are different for different classes of decision. Thus we might say that in baseball certain decisions, like the decision whether the ball or the runner reached second base first, are left to the discretion of the second base umpire, if we mean that on this issue the head umpire has no power to substitute his own judgment if he disagrees.

I call both of these senses weak to distinguish them from a stronger sense. We use 'discretion' sometimes not merely to say that an official must use judgment in applying the standards set him by authority, or that no one will review that exercise of judgment, but to say that on some issue he is simply not bound by standards set by the authority in question. In this sense we say that a sergeant has discretion who has been told to pick any five men for patrol he chooses or that a judge in a dog show has discretion to judge airedales before boxers if the rules do not stipulate an order of events. We use this sense not to comment on the vagueness or difficulty of the standards, or on who has the final word in applying them, but on their range and the decisions they purport to control. If the sergeant is told to take the five most experienced men, he does not have discretion in this strong sense because that order purports to govern his decision. The boxing referee who must decide which fighter has been the more aggressive does not have discretion, in the strong sense, for the same reason.¹

If anyone said that the sergeant or the referee had discretion in these cases, we should have to understand him, if the context permitted, as using the term in one of the weak senses. Suppose, for example, the

¹ I have not spoken of that jurisprudential favorite, 'limited' discretion, because that concept presents no special difficulties if we remember the relativity of discretion. Suppose the sergeant is told to choose from 'amongst' experienced men, or to 'take experience into account'. We might say either that he has (limited) discretion in picking his patrol, or (full) discretion to either pick amongst experienced men or decide what else to take into account.

lieutenant ordered the sergeant to select the five men he deemed most experienced, and then added that the sergeant had discretion to choose them. Or the rules provided that the referee should award the round to the more aggressive fighter, with discretion in selecting him. We should have to understand these statements in the second weak sense, as speaking to the question of review of the decision. The first weak sense – that the decisions take judgment – would be otiose, and the third, strong sense is excluded by the statements themselves.

We must avoid one tempting confusion. The strong sense of discretion is not tantamount to license, and does not exclude criticism. Almost any situation in which a person acts (including those in which there is no question of decision under special authority, and so no question of discretion) makes relevant certain standards of rationality, fairness, and effectiveness. We criticize each other's acts in terms of these standards, and there is no reason not to do so when the acts are within the center rather than beyond the perimeter of the doughnut of special authority. So we can say that the sergeant who was given discretion (in the strong sense) to pick a patrol did so stupidly or maliciously or carelessly, or that the judge who had discretion in the order of viewing dogs made a mistake because he took boxers first although there were only three airedales and many more boxers. An official's discretion means not that he is free to decide without recourse to standards of sense and fairness, but only that his decision is not controlled by a standard furnished by the particular authority we have in mind when we raise the question of discretion. Of course this latter sort of freedom is important; that is why we have the strong sense of discretion. Someone who has discretion in this third sense can be criticized, but not for being disobedient, as in the case of the soldier. He can be said to have made a mistake, but not to have deprived a participant of a decision to which he was entitled, as in the case of a sports official or contest judge.

We may now return, with these observations in hand, to the positivists' doctrine of judicial discretion. That doctrine argues that if a case is not controlled by an established rule, the judge must decide it by exercising discretion. We want to examine this doctrine and to test its bearing on our treatment of principles; but first we must ask in which sense of discretion we are to understand it.

Some nominalists argue that judges always have discretion, even when a clear rule is in point, because judges are ultimately the final arbiters of the law. This doctrine of discretion uses the second weak sense of that term, because it makes the point that no higher authority reviews the decisions of the highest court. It therefore has no bearing on the issue of how we account for principles, any more than it bears on how we account for rules.

The positivists do not mean their doctrine this way, because they say that a judge has no discretion when a clear and established rule is available. If we attend to the positivists' arguments for the doctrine we may suspect that they use discretion in the first weak sense to mean only that judges must sometimes exercise judgment in applying legal standards. Their arguments call attention to the fact that some rules of law are vague (Professor Hart, for example, says that all rules of law have 'open texture'), and that some cases arise (like *Henningsen*) in which no established rule seems to be suitable. They emphasize that judges must sometimes agonize over points of law, and that two equally trained and intelligent judges will often disagree.

These points are easily made; they are commonplace to anyone who has any familiarity with law. Indeed, that is the difficulty with assuming that positivists mean to use 'discretion' in this weak sense. The proposition that when no clear rule is available discretion in the sense of judgment must be used is a tautology. It has no bearing, moreover, on the problem of how to account for legal principles. It is perfectly consistent to say that the judge in *Riggs*, for example, had to use judgment, and that he was bound to follow the principle that no man may profit from his own wrong. The positivists speak as if their doctrine of judicial discretion is an insight rather than a tautology, and as if it does have a bearing on the treatment of principles. Hart, for example, says that when the judge's discretion is in play, we can no longer speak of his being bound by standards, but must speak rather of what standards he 'characteristically uses'.¹ Hart thinks that when judges have discretion, the principles they cite must be treated on our second approach, as what courts 'make it a principle' to do.

It therefore seems that positivists, at least sometimes, take their doctrine in the third, strong sense of discretion. In that sense it does bear on the treatment of principles; indeed, in that sense it is nothing less than a restatement of our second approach. It is the same thing to say that when a judge runs out of rules he has discretion, in the sense that he is not bound by any standards from the authority of law, as to say that the legal standards judges cite other than rules are not binding on them.

So we must examine the doctrine of judicial discretion in the strong sense. (I shall henceforth use the term 'discretion' in that sense.) Do the principles judges cited in cases like *Riggs* or *Henningsen* control their decisions, as the sergeant's orders to take the most experienced men or the referee's duty to choose the more aggressive fighter control the decisions of these officials? What arguments could a positivist supply to show that they do not?

¹ H. L. A. Hart, *The Concept of Law*, 144 (1961).

(1) A positivist might argue that principles cannot be binding or obligatory. That would be a mistake. It is always a question, of course, whether any particular principle is *in fact* binding upon some legal official. But there is nothing in the logical character of a principle that renders it incapable of binding him. Suppose that the judge in *Henningsen* had failed to take any account of the principle that automobile manufacturers have a special obligation to their consumers, or the principle that the courts seek to protect those whose bargaining position is weak, but had simply decided for the defendant by citing the principle of freedom of contract without more. His critics would not have been content to point out that he had not taken account of considerations that other judges have been attending to for some time. Most would have said that it was his duty to take the measure of these principles and that the plaintiff was entitled to have him do so. We mean no more, when we say that a *rule* is binding upon a judge, than that he must follow it if it applies, and that if he does not he will on that account have made a mistake.

It will not do to say that in a case like *Henningsen* the court is only 'morally' obligated to take particular principles into account, or that it is 'institutionally' obligated, or obligated as a matter of judicial 'craft', or something of that sort. The question will still remain why this type of obligation (whatever we call it) is different from the obligation that rules impose upon judges, and why it entitles us to say that principles and policies are not part of the law but are merely extra-legal standards 'courts characteristically use'.

(2) A positivist might argue that even though some principles are binding, in the sense that the judge must take them into account, they cannot determine a particular result. This is a harder argument to assess because it is not clear what it means for a standard to 'determine' a result. Perhaps it means that the standard *dictates* the result whenever it applies so that nothing else counts. If so, then it is certainly true that the individual principles do not determine results, but that is only another way of saying that principles are not rules. Only rules dictate results, come what may. When a contrary result has been reached, the rule has been abandoned or changed. Principles do not work that way; they incline a decision one way, though not conclusively, and they survive intact when they do not prevail. This seems no reason for concluding that judges who must reckon with principles have discretion because a set of principles *can* dictate a result. If a judge believes that principles he is bound to recognize point in one direction and that principles pointing in the other direction, if any, are not of equal weight, then he must decide accordingly, just as he must follow what he believes

to be a binding rule. He may, of course, be wrong in his assessment of the principles, but he may also be wrong in his judgment that the rule is binding. The sergeant and the referee, we might add, are often in the same boat. No one factor dictates which soldiers are the most experienced or which fighter the more aggressive. These officials must make judgments of the relative weights of these various factors; they do not on that account have discretion.

(3) A positivist might argue that principles cannot count as law because their authority, and even more so their weight, are congenitally *controversial*. It is true that generally we cannot *demonstrate* the authority or weight of a particular principle as we can sometimes demonstrate the validity of a rule by locating it in an act of Congress or in the opinion of an authoritative court. Instead, we make a case for a principle, and for its weight, by appealing to an amalgam of practice and other principles in which the implications of legislative and judicial history figure along with appeals to community practices and understandings. There is no litmus paper for testing the soundness of such a case – it is a matter of judgment, and reasonable men may disagree. But again this does not distinguish the judge from other officials who do not have discretion. The sergeant has no litmus paper for experience, the referee none for aggressiveness. Neither of these has discretion, because he is bound to reach an understanding, controversial or not, of what his orders or the rules require, and to act on that understanding. That is the judge's duty as well.

Of course, if the positivists are right in another of their doctrines – the theory that in each legal system there is an ultimate *test* for binding law like Professor Hart's rule of recognition – it follows that principles are not binding law. But the incompatibility of principles with the positivists' theory can hardly be taken as an argument that principles must be treated any particular way. That begs the question; we are interested in the status of principles because we want to evaluate the positivists' model. The positivist cannot defend his theory of a rule of recognition by fiat; if principles are not amenable to a test he must show some other reason why they cannot count as law. Since principles seem to play a role in arguments about legal obligation (witness, again, *Riggs* and *Henningsen*), a model that provides for that role has some initial advantage over one that excludes it, and the latter cannot properly be inveighed in its own support.

These are the most obvious of the arguments a positivist might use for the doctrine of discretion in the strong sense, and for the second approach to principles. I shall mention one strong counter-argument

against that doctrine and in favor of the first approach. Unless at least some principles are acknowledged to be binding upon judges, requiring them as a set to reach particular decisions, then no rules, or very few rules, can be said to be binding upon them either.

In most American jurisdictions, and now in England also, the higher courts not infrequently reject established rules. Common law rules – those developed by earlier court decisions – are sometimes overruled directly, and sometimes radically altered by further development. Statutory rules are subjected to interpretation and reinterpretation, sometimes even when the result is not to carry out what is called the 'legislative intent.'¹ If courts had discretion to change established rules, then these rules would of course not be binding upon them, and so would not be law on the positivists' model. The positivist must therefore argue that there are standards, themselves binding upon judges, that determine when a judge may overrule or alter an established rule, and when he may not.

When, then, is a judge permitted to change an existing rule of law? Principles figure in the answer in two ways. First, it is necessary, though not sufficient, that the judge find that the change would advance some principle, which principle thus justifies the change. In *Riggs* the change (a new interpretation of the statute of wills) was justified by the principle that no man should profit from his own wrong; in *Henningsen* the previously recognized rules about automobile manufacturers' liability were altered on the basis of the principles I quoted from the opinion of the court.

But not any principle will do to justify a change, or no rule would ever be safe. There must be some principles that count and others that do not, and there must be some principles that count for more than others. It could not depend on the judge's own preferences amongst a sea of respectable extra-legal standards, any one in principle eligible, because if that were the case we could not say that any rules were binding. We could always imagine a judge whose preferences amongst extra-legal standards were such as would justify a shift or radical reinterpretation of even the most entrenched rule.

Second, any judge who proposes to change existing doctrine must take account of some important standards that argue against departures from established doctrine, and these standards are also for the most part principles. They include the doctrine of 'legislative supremacy', a set of principles that require the courts to pay a qualified deference to the acts of the legislature. They also include the doctrine of precedent, another set of principles reflecting the equities and efficiencies of consistency. The doctrines of legislative supremacy and precedent incline

¹ See Wellington and Albert, 'Statutory Interpretation and the Political Process: A Comment on *Sinclair v. Atkinson*', 72 *Yale L. J.* 1547 (1963).

toward the *status quo*, each within its sphere, but they do not command it. Judges are not free, however, to pick and choose amongst the principles and policies that make up these doctrines – if they were, again, no rule could be said to be binding.

Consider, therefore, what someone implies who says that a particular rule is binding. He may imply that the rule is affirmatively supported by principles the court is not free to disregard, and which are collectively more weighty than other principles that argue for a change. If not, he implies that any change would be condemned by a combination of conservative principles of legislative supremacy and precedent that the court is not free to ignore. Very often, he will imply both, for the conservative principles, being principles and not rules, are usually not powerful enough to save a common law rule or an aging statute that is entirely unsupported by substantive principles the court is bound to respect. Either of these implications, of course, treats a body of principles and policies as law in the sense that rules are; it treats them as standards binding upon the officials of a community, controlling their decisions of legal right and obligation.

We are left with this issue. If the positivists' theory of judicial discretion is either trivial because it uses 'discretion' in a weak sense, or unsupported because the various arguments we can supply in its defense fall short, why have so many careful and intelligent lawyers embraced it? We can have no confidence in our treatment of that theory unless we can deal with that question. It is not enough to note (although perhaps it contributes to the explanation) that 'discretion' has different senses that may be confused. We do not confuse these senses when we are not thinking about law.

Part of the explanation, at least, lies in a lawyer's natural tendency to associate laws and rules, and to think of 'the law' as a collection or system of rules. Roscoe Pound, who diagnosed this tendency long ago, though that English speaking lawyers were tricked into it by the fact that English uses the same word, changing only the article, for 'a law' and 'the law'.¹ (Other languages, on the contrary, use two words: 'loi' and 'droit', for example, and 'Gesetz' and 'Recht'.) This may have had its effect, with the English speaking positivists, because the expression 'a law' certainly does suggest a rule. But the principal reason for associating law with rules runs deeper, and lies, I think, in the fact that legal education has for a long time consisted of teaching and examining those established rules that form the cutting edge of law.

In any event, if a lawyer thinks of law as a system of rules, and yet recognizes, as he must, that judges change old rules and introduce new

ones, he will come naturally to the theory of judicial discretion in the strong sense. In those other systems of rules with which he has experience (like games), the rules are the only special authority that govern official decisions, so that if an umpire could change a rule, he would have discretion as to the subject matter of that rule. Any principles umpires might mention when changing the rules would represent only their 'characteristic' preferences. Positivists treat law like baseball revised in this way.

There is another, more subtle consequence of this initial assumption that law is a system of rules. When the positivists do attend to principles and policies, they treat them as rules *manquées*. They assume that if they are standards of law they must be rules, and so they read them as standards that are trying to be rules. When a positivist hears someone argue that legal principles are part of the law, he understands this to be an argument for what he calls the 'higher law' theory, that these principles are the rules of a law about the law.¹ He refutes this theory by pointing out that these 'rules' are sometimes followed and sometimes not, that for every 'rule' like 'no man shall profit from his own wrong' there is another competing 'rule' like 'the law favors security of title', and that there is no way to test the validity of 'rules' like these. He concludes that these principles and policies are not valid rules of a law above the law, which is true, because they are not rules at all. He also concludes that they are extra-legal standards which each judge selects according to his own lights in the exercise of his discretion, which is false. It is as if a zoologist had proved that fish are not mammals, and then concluded that they are really only plants.

6. THE RULE OF RECOGNITION

This discussion was provoked by our two competing accounts of legal principles. We have been exploring the second account, which the positivists seem to adopt through their doctrine of judicial discretion, and we have discovered grave difficulties. It is time to return to the fork in the road. What if we adopt the first approach? What would the consequences of this be for the skeletal structure of positivism? Of course we should have to drop the second tenet, the doctrine of judicial discretion (or, in the alternative, to make plain that the doctrine is to be read merely to say that judges must often exercise judgment). Would we also have to abandon or modify the first tenet, the proposition that law is distinguished by tests of the sort that can be set out in a master rule like Professor Hart's rule of recognition? If principles of the *Riggs* and

¹ R. Pound, *An Introduction to the Philosophy of Law* 56 (rev. ed. 1954).

¹ See, e.g., Dickinson, 'The Law Behind Law (pts. 1 & 2)', 29, *Columbia Law Review* 112, 254 (1929).

Henningsen sort are to count as law, and we are nevertheless to preserve the notion of a master rule for law, then we must be able to deploy some test that all (and only) the principles that do count as law meet. Let us begin with the test Hart suggests for identifying valid *rules* of law, to see whether these can be made to work for principles as well.

Most rules of law, according to Hart, are valid because some competent institution enacted them. Some were created by a legislature, in the form of statutory enactments. Others were created by judges who formulated them to decide particular cases, and thus established them as precedents for the future. But this test of pedigree will not work for the *Riggs* and *Henningsen* principles. The origin of these as legal principles lies not in a particular decision of some legislature or court, but in a sense of appropriateness developed in the profession and the public over time. Their continued power depends upon this sense of appropriateness being sustained. If it no longer seemed unfair to allow people to profit by their wrongs, or fair to place special burdens upon oligopolies that manufacture potentially dangerous machines, these principles would no longer play much of a role in new cases, even if they had never been overruled or repealed. (Indeed, it hardly makes sense to speak of principles like these as being 'overruled' or 'repealed'. When they decline they are eroded, not torpedoed.)

True, if we were challenged to back up our claim that some principle is a principle of law, we would mention any prior cases in which that principle was cited, or figured in the argument. We would also mention any statute that seemed to exemplify that principle (even better if the principle was cited in the preamble of the statute, or in the committee reports or other legislative documents that accompanied it). Unless we could find some such institutional support, we would probably fail to make out our case, and the more support we found, the more weight we could claim for the principle.

Yet we could not devise any formula for testing how much and what kind of institutional support is necessary to make a principle a legal principle, still less to fix its weight at a particular order of magnitude. We argue for a particular principle by grappling with a whole set of shifting, developing and interacting standards (themselves principles rather than rules) about institutional responsibility, statutory interpretation, the persuasive force of various sorts of precedent, the relation of all these to contemporary moral practices, and hosts of other such standards. We could not bolt all of these together into a single 'rule', even a complex one, and if we could the result would bear little relation to Hart's picture of a rule of recognition, which is the picture of a fairly stable master rule specifying 'some feature or features possession of which by

a suggested rule is taken as a conclusive affirmative indication that it is a rule. . .¹

Moreover, the techniques we apply in arguing for another principle do not stand (as Hart's rule of recognition is designed to) on an entirely different level from the principles they support. Hart's sharp distinction between acceptance and validity does not hold. If we are arguing for the principle that a man should not profit from his own wrong, we could cite the acts of courts and legislatures that exemplify it, but this speaks as much to the principle's acceptance as its validity. (It seems odd to speak of a principle as being valid at all, perhaps because validity is an all-or-nothing concept, appropriate for rules, but inconsistent with a principle's dimension of weight.) If we are asked (as we might well be) to defend the particular doctrine of precedent, or the particular technique of statutory interpretation, that we used in this argument, we should certainly cite the practice of others in using that doctrine or technique. But we should also cite other general principles that we believe support that practice, and this introduces a note of validity into the chord of acceptance. We might argue, for example, that the use we make of earlier cases and statutes is supported by a particular analysis of the point of the practice of legislation or the doctrine of precedent, or by the principles of democratic theory, or by a particular position on the proper division of authority between national and local institutions, or something else of that sort. Nor is this path of support a one-way street leading to some ultimate principle resting on acceptance alone. Our principles of legislation, precedent, democracy, or federalism might be challenged too; and if they were we should argue for them, not only in terms of practice, but in terms of each other and in terms of the implications of trends of judicial and legislative decisions, even though this last would involve appealing to those same doctrines of interpretation we justified through the principles we are now trying to support. At this level of abstraction, in other words, principles rather hang together than link together.

So even though principles draw support from the official acts of legal institutions, they do not have a simple or direct enough connection with these acts to frame that connection in terms of criteria specified by some ultimate master rule of recognition. Is there any other route by which principles might be brought under such a rule?

Hart does say that a master rule might designate as law not only rules enacted by particular legal institutions, but rules established by *custom* as well. He has in mind a problem that bothered other positivists, including Austin. Many of our most ancient legal rules were never

¹ H. L. A. Hart, *The Concept of Law* 92 (1961).

explicitly created by a legislature or a court. When they made their first appearance in legal opinions and texts, they were treated as already being part of the law because they represented the customary practice of the community, or some specialized part of it, like the business community. (The examples ordinarily given are rules of mercantile practice, like the rules governing what rights arise under a standard form of commercial paper.)¹ Since Austin thought that all law was the command of a determinate sovereign, he held that these customary practices were not law until the courts (as agents of the sovereign) recognized them, and that the courts were indulging in a fiction in pretending otherwise. But that seemed arbitrary. If everyone thought custom might in itself be law, the fact that Austin's theory said otherwise was not persuasive.

Hart reversed Austin on this point. The master rule, he says, might stipulate that some custom counts as law even before the courts recognize it. But he does not face the difficulty this raises for his general theory because he does not attempt to set out the criteria a master rule might use for this purpose. It cannot use, as its only criterion, the provision that the community regard the practice as *morally* binding, for this would not distinguish legal customary rules from moral customary rules, and of course not all of the community's long-standing customary moral obligations are enforced at law. If, on the other hand, the test is whether the community regards the customary practice as *legally* binding, the whole point of the master rule is undercut, at least for this class of legal rules. The master rule, says Hart, marks the transformation from a primitive society to one with law, because it provides a test for determining social rules of law other than by measuring their acceptance. But if the master rule says merely that whatever other rules the community accepts as legally binding are legally binding, then it provides no such test at all, beyond the test we should use were there no master rule. The master rule becomes (for these cases) a non-rule of recognition; we might as well say that every primitive society has a secondary rule of recognition, namely the rule that whatever is accepted as binding is binding. Hart himself, in discussing international law, ridicules the idea that such a rule could be a rule of recognition, by describing the proposed rule as 'an empty repetition of the mere fact that the society concerned . . . observes certain standards of conduct as obligatory rules'.²

¹ See Note, 'Custom and Trade Usage: Its Application to Commercial Dealings and the Common Law', 55 *Columbia Law Review* 1192 (1955), and materials cited therein at 1193 n.1. As that note makes plain, the actual practices of courts in recognizing trade customs follow the pattern of applying a set of general principles and policies rather than a test that could be captured as part of a rule of recognition.

² H. L. Hart, *The Concept of Law* 230 (1961). A master rule might specify some particular feature of a custom that is independent of the community's attitude; it

Hart's treatment of custom amounts, indeed, to a confession that there are at least some rules of law that are not binding because they are valid under standards laid down by a master rule but are binding – like the master rule – because they are accepted as binding by the community. This chips at the neat pyramidal architecture we admired in Hart's theory: we can no longer say that only the master rule is binding because of its acceptance, all other rules being valid under its terms.

This is perhaps only a chip, because the customary rules Hart has in mind are no longer a very significant part of the law. But it does suggest that Hart would be reluctant to widen the damage by bringing under the head of 'custom' all those crucial principles and policies we have been discussing. If he were to call these part of the law and yet admit that the only test of their force lies in the degree to which they are accepted as law by the community or some part thereof, he would very sharply reduce that area of the law over which his master rule held any dominion. It is not just that all the principles and policies would escape its sway, though that would be bad enough. Once these principles and policies are accepted as law, and thus as standards judges must follow in determining legal obligations, it would follow that *rules* like those announced for the first time in *Riggs* and *Henningsen* owe their force at least in part to the authority of principles and policies, and so not entirely to the master rule of recognition.

So we cannot adapt Hart's version of positivism by modifying his rule of recognition to embrace principles. No tests of pedigree, relating principles to acts of legislation, can be formulated, nor can his concept of customary law, itself an exception to the first tenet of positivism, be made to serve without abandoning that tenet altogether. One more possibility must be considered, however. If no rule of recognition can provide a test for identifying principles, why not say that principles are ultimate, and *form* the rule of recognition of our law? The answer to the general question 'What is valid law in an American jurisdiction?' would then require us to state all the principles (as well as ultimate constitutional rules) in force in that jurisdiction at the time, together with appropriate assignments of weight. A positivist might then regard the complete set of these standards as the rule of recognition of the jurisdiction. This

might provide, for example, that all customs of very great age, or all customs having to do with negotiable instruments count as law. I can think of no such features that in fact distinguish the customs that have been recognized as law in England or America, however. Some customs that are not legally enforceable are older than some that are, some practices relating to commercial paper are enforced and others not, and so forth. In any event, even if a distinguishing feature were found that identified all rules of law established by custom, it would remain unlikely that such a feature could be found for principles which vary widely in their subject matter and pedigree and some of which are of very recent origin.

solution has the attraction of paradox, but of course it is an unconditional surrender. If we simply designate our rule of recognition by the phrase 'the complete set of principles in force', we achieve only the tautology that law is law. If, instead, we tried actually to list all the principles in force we would fail. They are controversial, their weight is all important, they are numberless, and they shift and change so fast that the start of our list would be obsolete before we reached the middle. Even if we succeeded, we would not have a key for law because there would be nothing left for our key to unlock.

I conclude that if we treat principles as law we must reject the positivists' first tenet, that the law of a community is distinguished from other social standards by some test in the form of a master rule. We have already decided that we must then abandon the second tenet – the doctrine of judicial discretion – or clarify it into triviality. What of the third tenet, the positivists' theory of legal obligation?

This theory holds that a legal obligation exists when (and only when) an established rule of law imposes such an obligation. It follows from this that in a hard case – when no such established rule can be found – there is no legal obligation until the judge creates a new rule for the future. The judge may apply that new rule to the parties in the case, but this is *ex post facto* legislation, not the enforcement of an existing obligation.

The positivists' doctrine of discretion (in the strong sense) required this view of legal obligation, because if a judge has discretion there can be no legal right or obligation – no entitlement – that he must enforce. Once we abandon that doctrine, however, and treat principles as law, we raise the possibility that a legal obligation might be imposed by a constellation of principles as well as by an established rule. We might want to say that a legal obligation exists whenever the case supporting such an obligation, in terms of binding legal principles of different sorts, is stronger than the case against it.

Of course, many questions would have to be answered before we could accept that view of legal obligation. If there is no rule of recognition, no test for law in that sense, how do we decide which principles are to count, and how much, in making such a case? How do we decide whether one case is better than another? If legal obligation rests on an undemonstrable judgment of that sort, how can it provide a justification for a judicial decision that one party had a legal obligation? Does this view of obligation square with the way lawyers, judges and laymen speak, and is it consistent with our attitudes about moral obligation? Does this analysis help us to deal with the classical jurisprudential puzzles about the nature of law?

These questions must be faced, but even the questions promise more

than positivism provides. Positivism, on its own thesis, stops short of just those puzzling, hard cases that send us to look for theories of law. When we read these cases, the positivist remits us to a doctrine of discretion that leads nowhere and tells nothing. His picture of law as a system of rules has exercised a tenacious hold on our imagination, perhaps through its very simplicity. If we shake ourselves loose from this model of rules, we may be able to build a model truer to the complexity and sophistication of our own practices.