Modeling Legal Rules

RICHARD HOLTON, MIT

Common law rules admit of exceptions. When a court, especially a higher court, finds that the routine application of a rule would result in an injustice it is likely to distinguish: to concede that yes, the case does appear to fall under the rule as it is currently understood, but to insist that there are further factors, not mentioned in the rule (though perhaps acknowledged in other rules in other parts of the law) that distinguish this case from the cases that the existing rule was meant to cover, and that mean that the verdict that the existing rule suggests would be wrong. Nevertheless, the old rule does not die.[1] When the writers of case books come to accommodate the new ruling it comes in as an amendment: the old rule was correct except under these new circumstances.

So how should we understand the form of legal rules? A simple minded approach is to see them as universally quantified claims: whenever this holds, then this is the right verdict. But exceptions make it hard to maintain this. A universally quantified sentence cannot have exceptions, only counterexamples; and counterexamples show that the sentence is false. Of course, one could insist on the approach, maintaining that a simple legal rule is indeed false, and that the need to amend it shows this to be so. But since every common law rule, however amended, is very likely to admit of further amendment, this leaves us in the uncomfortable position of saying no rule is strictly true. We might try to soften the blow by saying that as they are amended the rules get closer to the truth, but proposals to explain such an idea have been signally unsuccessful.[2]

So we need an approach that allows rules to have exceptions. Elsewhere I have developed an account that sees legal rules as universals containing implicit *unless*-clauses; the idea is that the exceptions trigger the clauses.[3] The challenge comes in doing this in such a way that the rules do not become trivial. In the first part of the paper I summarize that account (and correct what I now take to be some errors in my initial formulation). In the second part of the paper I ask how it fares against some alternatives. I see two. One treats legal rules not as universal generalizations at all, but as generics. I argue that whilst this has some plausibility for legal principles, it doesn't do the job for legal rules. The second alternative is more radical still, treating legal rules as default rules within a nonmonotonic logic. Here I argue that the move to nonmonotonic logic does not really bring the advantages claimed for it, and further that it fails to explain something that is handled very nicely by the approach I favour: how it is that a legal decision can be criticizable, even though the court used the legal rules that were in force at the time.

---

[1] Or at least, hardly ever. When the Australian High Court gave the *Mabo* decision, possibly the whole set of rules based around terra nullius were thrown out. But that is unusual.

[2] This was part of Karl Popper's idea of verisimilitude. See David Miller and Pavel Tichy for the problems with it.

[3] See 'The Exception Proves the Rule', *Journal of Political Philosophy* forthcoming; available on my website: http://web.mit.edu/holton/www/pubs.html

Particularists in ethics hold that any rule will be subject to exceptions.[4] Let us concede them the point: let us accept that however many qualifications we insert into a moral rule, we will still be able to find further counterexamples. For all we know, that is right; certainly we have some good inductive grounds for it. Particularists then typically conclude that there are no true moral rules. Here I differ. I am in good company. Hart writes:

> We promise to visit a friend the next day. When the day comes it turns out that keeping the promise would involve neglecting someone dangerously ill. The fact that this is accepted as an adequate reason for not keeping the promise surely does not mean that there is no rule requiring promises to be kept, only a certain regularity in keeping them. It does not follow from the fact that rules have exceptions incapable of exhaustive statement, that in every situation we are left to our discretion and are never bound to keep a promise. A rule that ends with the word 'unless . . .' is still a rule. [5]

So Hart at least implicitly accepts that some valid rules have exceptions that cannot be exhaustively stated. His comments come after his famous discussion of the open texture of law that results from the open texture of language—Is a child's bicycle covered by a regulation that prohibits vehicles from the park?—so one might think that the ideas here can be similarly explained. But clearly they cannot. Whilst there may be vagueness in the idea of a promise—Is it a promise if made under duress, or if the promisor does not understand what they are committing themselves to, or if the promisee is not aware of it?—such vagueness is not relevant here. There may be no doubt that my promise was as clear and central an example as one is ever likely to find, and yet it still be true that in the circumstances I am not bound to keep it. So we need a different explanation.

One complicating factor is that Hart's discussion here concerns moral rules and obligations, rather than legal. This is surprising, since Hart has been talking about legal rules, and it is far from obvious that what holds for moral rules also holds for legal. We'll shortly return to the question of whether the two should be treated in the same way; for now let us follow Hart and keep our focus on the moral.

Hart's example might put us in mind of recent discussions of moral particularism. Those who advocate such an approach contend that any putative moral rule is subject to exceptions; and as a result they tend to reject any role for rules. Take any rule that links the moral to the descriptive, they say, and we can find an exception to it; amend the rule to embrace the exception and we can find an exception to the amended rule, and so on.

I am sympathetic to the drift of the position, but the as characterized it strikes me as too strong on several counts. First, how could we be sure that *every* moral rule is subject to exception? Our best grounds are inductive: that despite much effort, every rule that has been proposed has been shown vulnerable to exception. But it is a long and dogmatic step

---

[4] Jonathan Dancy, 'Ethical Particularism and Morally Relevant Properties' *Mind* 92 (1983) 530–47; *Ethics Without Principles* (Oxford: Clarendon Press 2004).

[5] H.L.A. Hart, *Concept of Law* (Oxford: Clarendon Press, 1961) p. 136

from that to the insistence that there are no exceptionless rules. A more plausible claim is a more modest one. Particularism should be couched in terms of its opposition to the theorist on the other side who insists that there are exceptionless rules. Given the failure of previous attempts to find them, the insistence on the existence of such rules is the piece of dogmatism. The particularist position should then be characterized as scepticism about that position, coupled with the positive claim that in the light of such scepticism we should not construct our moral theories on the assumption that exceptionless rules are there to be found.

Second, it does not follow from the supposition that there are no exceptionless moral rules that rules have no role to play. As we have seen, Hart suggests a model: perhaps rules end with an (often unstated) *unless*-clause; where this is not triggered, the rule applies. Hart, however, does not tell us how the *unless*-clause is to be completed. And here it might seem that he is faced with a dilemma. On one approach the *unless*-clause contains a full statement of all the factors that would defeat the rule; but that is clearly incompatible with the idea that the exceptions are incapable of exhaustive statement. On the other approach the *unless*-clause would be open ended: 'One should keep one's promises unless there is reason not to'. But that is clearly trivial. If we are to give substance to an open-ended *unless*-clause we need to find a middle way between these two approaches. I think that something is available for moral rules. In the next section I'll present this account. Then I shall turn to the question of whether it is applicable to the law.

MORALS

The intuitive idea that I shall work with is that a moral rule can be over-ruled if there is a moral justification for the exception.[6] In Hart's example there are moral grounds for tending to the seriously ill person rather than keeping my promise. But if so, then there is plausibly a moral rule that tells us that there are such grounds. So the *unless*-clause can be read as quantifying over other moral rules. It says that the moral rule will apply to the case unless there are other moral rules that apply to that case that render the verdict of the first rule wrong. That is: a moral rule like "Killing is wrong" applies to a case of killing if and only if there are no other moral rules—for instance, "Killing in self-defence is not wrong"—that apply to the case and render the verdict of the first rule wrong.[7] But the *unless*-clause doesn't list all of the possible further rules that would defeat the application of the initial rule. That would be impossible if they are incapable of exhaustive statement as Hart supposes. It simply quantifies over them.

The crucial thought here is that what makes a moral rule apply to a case isn't just what obtains; it is also what doesn't obtain. So as well as adding an *unless*-clause to the moral

---

[6] I here follow the account that I gave in 'Principles and Particularisms' *Proceedings of the Aristotelian Society Supplementary Volume* 67, (2002) pp. 191-209. Readers wanting more details of the account, and further discussion of difficulties it faces, should consult that article. Note that there I used the term 'principle' for what I am here calling a rule. Following Dworkin I now reserve 'principle' for something couched at a greater level of generality; see below

[7] Phrased like this, perhaps the rules are generics, not properly analyzed a universally quantified generalizations at all. I discuss this below. If that is right, then the move to the formalization is not simply translation; the formalization actually says more. I try to say what that might be below.

rules, the full form of a moral argument will also require the addition of a premise to the effect that the *unless*-clause is not triggered. We can make these ideas more precise by defining a notion of what it is for one set of considerations to be *superseded* by a second; that is, for the second set to allude to some further consideration that would upset the conclusion that one would reach on the basis of the first. Then we can construct the *unless*-clause, which I call '*That's it*', in terms of that notion of supersession.

I start with a first definition of supersession, one which will have to be modified, but which will serve to fix the main idea:

> *Supersession (first attempt)*
> Suppose we have a set of predicates $\{F_1, F_2 \ldots F_m\}$; and suppose that these occur in a set of sentences $\{F_1a, F_2a, \ldots F_m a\}$ and in a corresponding moral rule of the form $\forall x\, ((F_1 x$ & $F_2 x$ & $\ldots$ & $F_m x) \rightarrow F_v x)$, where F_v is a predicate expressing a moral verdict. Then we say that those sentences and that moral rule are *superseded* by another set of sentences $\{G_1a, G_2a, \ldots G_n a\}$ and a corresponding moral rule $\forall x\, ((G_1 x$ & $G_2 x$ & $\ldots$ & $G_n x) \rightarrow G_v x)$ if and only if:
>
>   (i) $(G_1 x$ & $G_2 x$ & $\ldots$ & $G_n x)$ entails $(F_1 x$ & $F_2 x$ & $\ldots$ & $F_m x)$, but not *vice versa;*
>   (ii) $F_v x$ is incompatible with $G_v x$.

The first clause here requires that the second set of set of sentences says everything that is said by the first *and* something more; the second clause requires that the second moral rule brings one to a verdict that is incompatible with that of the first. For instance, the sentence 'This was a killing', and the moral rule '$\forall x$ If x is a killing then x is wrong' are superseded by the sentences 'This was a killing' and 'This was done in self defence' and the rule '$\forall x$ If x is a killing and x is done in self defence then x is not wrong'. Now we can go on to define the *unless*-clause, making use of the idea of supersession that we have defined:

> *That's it*:   There are no further relevant moral rules and facts; i.e. there is no true moral rule and set of *true* sentences that supersede those that appear in this argument.

What we want to say is that if *That's it* holds, the moral rule holds. So we need to add *That's it* to the rule, as a further conjunct of the antecedent: 'If x is a killing, and *That's it*, then x is wrong'. Then the arguments in which the rule occurs need a further premise saying that *That's it* does indeed obtain.

So we get arguments like this:

Im	P1	This is a killing
	P2	$\forall x\, ((x \text{ is a killing}$ & *That's it*$) \rightarrow \text{you may not do x})$
	P3	*That's it*
	C	You may not do this.

Recall that Hart claimed that the *unless*-clause cannot be exhaustively stated: no matter how many exceptions are give to a rule, one can always imagine further exceptions that have not been captured. Equivalently, no matter how much is built into the content of the rule itself, one can always imagine further factors that will render the rule invalid. In our current framework, this can now be understood as the claim that any moral argument like MI is bound to be superseded by other valid arguments: take any moral argument, we can always find another that supersedes it. [8] Or, at least, that is what we want to say; unfortunately we can't quite say it yet, since the result of inserting the *That's it* clause into the arguments is that the original definition of supersession no longer applies to them. So let us stop to fix that up by redefining supersession:

> *Supersession (second attempt)*
> A set of sentences $\{F_1a, F_2a, \ldots F_ma\}$ and a corresponding moral rule $\forall x\,((F_1x\ \&\ F_2x\ \&\ \ldots\ \&\ F_mx\ \&\ \textit{That's it}) \rightarrow F_cx)$ are *superseded* by another set of sentences $\{G_1a, G_2a, \ldots G_na\}$ and a corresponding moral rule $\forall x\,((G_1x\ \&\ G_2x\ \&\ \ldots\ \&\ G_nx\ \&\ \textit{That's it}) \rightarrow G_cx)$ if and only if:
>
> (i) $(G_1x\ \&\ G_2x\ \&\ \ldots\ \&G_nx)$ entails $(F_1x\ \&\ F_2x\ \&\ \ldots\ \&\ F_mx)$, but not *vice versa*;
> (ii) F_cx is incompatible with G_cx.[9]

Given this revised definition we can say that IM is superseded by the valid argument:

2M PI This is a killing
P2 This is done in self defence
P2 $\forall x\,((x$ is a killing $\&\ x$ is done in self defence $\&\ \textit{That's it}) \rightarrow$ you may do $x)$
P3 *That's it*
C You may do this.

Similarly 2M would be superseded by an argument that added the claim that the killing was not necessary for the defence, and that in turn would superseded by one that added that the defendant didn't realize this to be so, and that by one that he could have known it had he only paid due care, and so on. If Hart is right, no matter how complicated the rule gets, we will always be able to think of an argument that supersedes it. But the fact that every moral argument is superseded by some *valid* argument does not mean that it is superseded by a *sound* argument, that is, by a valid argument that has *true* premises. If the killing was not done is self-defence, and there is equally no other excusing condition, then the *That's it* premise in the original argument IM will be true, IM will not be superseded by any sound argument, and the conclusion, that the killing was impermissible, will be simply true. Put another way: the fact that every moral argument would be superseded

---

[8] I say that one *argument* supersedes a second iff the premises of the first argument, with the *That's it* content removed (both the premise, and the conjunct in the rule), supersede the premises of the second argument, with its *That's it* content removed. Since *That's it* includes a demonstrative, it changes its sense when it occurs in two different arguments, so we cannot define a notion of supersession that includes the *That's it* material.

[9] One consequence of this is that the definitions of supersession and *That's it* are now circular, in the sense that each makes reference to the other. I don't think that this is pernicious. For discussion see 'Principles and Particularisms' p. 200.

were certain facts to obtain does not show that every moral argument is in fact superseded. A good moral argument is one that is not.

So the approach meets one of the desiderata with which we started: we have found a way to interpret the *unless*-clause that does not involve a simple list. The approach also meets the second, for clearly it does not lead to triviality. The *That's it* condition is a substantial one. Many real moral arguments go wrong exactly because it is not met. Very often when we reach a mistaken moral conclusion it is because there is some further relevant factor that we are overlooking.

THE LAW

Can we take a similar approach in the legal case? Formally this is straightforward. We need a parallel account of supersession, defined as we defined it in the moral case, except that talk of moral rules is replaced with talk of legal rules. Thus, for instance, the legal rule '∀x If x dishonestly appropriates property belonging to another with the intention of permanently depriving the other of it, and *That's it*, then x is guilty of theft' and the sentence 'A dishonestly appropriated property belonging to another with the intention of permanently depriving the other of it' are superseded by the legal rule '∀x If x dishonestly appropriates property belonging to another with the intention of permanently depriving the other of it, and x is coerced, and That's it, then x is not guilty of theft' and the sentences 'A dishonestly appropriated property belonging to another with the intention of permanently depriving the other of it' and 'A was coerced'.

Then we need a version of *That's it* that applies to legal rules:

That's it: There are no further relevant facts and legal rules; i.e. there is no set of true sentences that, together with a legal rule, would supersede those that appear in this argument

Now we get arguments like this:

1L P1 Jones dishonestly appropriated property belonging to another with the intention of permanently depriving the other of it

P2 ∀x ((If x dishonestly appropriated property belonging to another with the intention of permanently depriving the other of it & *That's it*) → x is guilty of theft)

P3 *That's it*

C Jones is guilty of theft

The claim is that this argument could be superseded by other *valid* arguments, for instance

2L P1 Jones dishonestly appropriated property belonging to another with the intention of permanently depriving the other of it

P2 Jones acted under duress

P3   ∀x ((If x dishonestly appropriated property belonging to another with the intention of permanently depriving the other of it & x acted under duress & *That's it*) → x is not guilty of theft)

P4   *That's it*

C   Jones is not guilty of theft

But again the fact that every legal argument can be superseded by valid arguments does not mean that every legal argument is superseded by sound arguments. A good legal argument in a particular judgment is one in which the *That's it* clause is true.[10]

DISANALOGIES?

It is easy enough then to construct an account of legal argument that is formally analogous to that I gave for moral arguments. What is harder is to see whether the substance is the same. There is one glaring difference.[11] Moral particularism is typically associated with a form of objectivism about morals. On this approach it is exactly because morality is independent of us that it is uncodifiable. Though the picture is probably not compulsory, it is natural to imagine the moral truths there, ranged in infinite ranks ready to supersede any codification that we attempt.

There are views on which the law is similar: natural law theories that treat the law as not exhausted by the practices of the courts. But such views are contentious, and they are clearly not available to positivists like Hart, who see law as, in some sense, an artifact. Can the account be made compatible with a positivist approach?

The central issue is how we should understand the reference to a 'legal rule' in the definition of supersession and the *That's it* clause. It might seem that to a positivist like Hart the list of legitimate legal rules must be finite. After all, they are created as a result of human actions, and there has been a finite number of them. So if each action can itself create only a finite number of rules, then the number of rules must be finite. And one might think that this means our *unless*-clause will not be uncodifiable after all, and it will not be true that each legal argument will be potentially superseded.

One response here would be to soften the positivism. We might, for instance, broaden the account of what the law is, so as to include not just the extant sources, but also what follows from those sources, or what provides the best justification for those sources. Alternatively, leaving the law as it is, we might want to soften the definition of supersession and the *That's it* clause, so that the relevant legal rules wouldn't be the extant

---

[10] Point of clarification: I am primarily talking here about exceptions that involve a change to the legal rule defining what is necessary for an offence. There are other cases in which an exception will involve conceding that the law has been broken, but will provide a full or partial defence given the nature of the action. I would suggest treating these in like way, although here the legal rule defining the offence would not be qualified; the qualification would rather apply to the further rule defining the appropriate treatment of the defendant.

[11] Perhaps there are other differences that tell in favour of the account for law rather than for morals. Jonathan Dancy objects that the account can make no sense of contributory reasons: *Ethics Without Principles*, 27–9. I'm not sure what to make of that worry in the moral case; in the legal case, given the role I assign to legal rules, I doubt that we need any such notion.

rules, but rather those that would be included in a legitimate extension of the extant rules, where again this might make use of some notion of inference from, or justification of, the extant rules.

There are things to be said for and against these approaches.[12] I shall not, however, pursue the matter here, for neither, I think, is necessary. That is, there is an important sense in which we can think of the law as being uncodifiable even if we stick with a restrictive, positivist, source-based account of what the law is, and think that the definition of supersession and the *That's it* clause quantifies over only such law. The trick involves thinking of the law *from the judge's perspective*.

When a judge decides a case, the law, at the moment of deciding that case, will include all of the previous law, *and* the very decision that the judge is making. If the judge has distinguished the case from those that went before, the decision will itself constitute a new legal rule. Suppose, to take an absurdly simplified example, that the law on theft had never confronted a case of coerced theft, so that every legal decision looked liked 1L. Then suppose a judge did confront such a case, and distinguished it from those covered by the existing legal rule to produce an argument like 2L. It would then be true at the time that the decision was made that there was a legal rule that superseded the earlier simple rule, since it would be the rule that appeared in that very decision. Moreover, of course, the antecedent of that rule would be true, so the argument would be sound.

If the possible decisions, even the possible legitimate decisions, that a judge could make are uncodifiable, then, when she comes to list the rules that she might use in making a decision, that list would be uncodifiable. Why might the possible legitimate decisions be uncodifiable? The obvious explanation is provided by the uncodifiability of morals. If, as a matter of fact, the legal rules contain moral requirements as soft positivists like Hart allow[13]—if, for instance they require that a given expectation be *reasonable*, or a burden not be *unfair*, or that there be *due* care—and if the application of those moral requirements is uncodifiable as the moral particularists allege, then the uncodifiability will be inherited by the legal rules.

I described this move as a 'trick' but I didn't mean that in a derogatory sense. There is no real trickery here. We are not saying that the legal rules at any one moment are uncodifiable; an observer could take a legal snapshot of how things stand, and could exhaustively characterize the content of the law, though of course this content might include moral terms that are themselves uncodifiable. We are rather saying that the possible legal rules confronting a judge at the moment of making a decision are

---

[12] For arguments against the first, see J Raz, 'Authority, Law and Morality' *The Monist*, vol. 68, 295-324; similar arguments could be applied to the second. My own primary worry with either approach stems from a scepticism that they would result in a unique consistent extension.

[13] There is a weaker and a stronger thesis that soft, or inclusive, positivists might accept. The weaker is the idea that certain moral requirements may have entered into the law by precedent; the stronger is the thesis that the rule of recognition might allow moral requirements to be incorporated into the law directly, with no basis in precedent. I need only the weaker thesis here. (For a nice presentation of the distinction see Brian Leiter, 'Beyond the Hart/Dworkin Debate: The Methodology Problem in Jurisprudence', *American Journal of Jurisprudence* 48 (2003). He takes the Hart/Raz dispute to concern only the stronger thesis. It would certainly suit me if the weaker thesis were uncontroversial, but it frequently seems as though some of those on the Raz side of the debate were denying it.)

uncodifiable; and since these rules would be made actual if the judge endorsed them, something that it is in her power to do, it is the uncodifiability of the possible rules that matters. The point is simply that, once we concede the degree of discretion to judges that is needed if they are to distinguish cases, then judges cannot regard the law as fixed at the point at which they decide a case.

We must not confuse this sense of discretion with another notion of discretion that is employed by Hart, a notion that is sometimes called 'strong discretion'.[14] This is the idea that in certain cases the law gives no right answer: judges can legitimately decide either way. Hart accepts that are instances of strong discretion—cases when the law simply leaves a gap—but he wants to greatly limit their extent. As we saw, he wants to say that rules can contain *unless*-clauses, and yet, when these clauses are not triggered, they will be binding.

Now though we might worry whether this position is really available to Hart. If every legal rule contains an *unless*-clause, and if judges have the discretion needed to distinguish exceptions, then what is to stop them at any point deciding to create a new rule that supersedes the rule that otherwise would have applied? Of course there might be a rule governing when an exception can be made; but if that rule in turn contains an *unless*-clause, that can be superseded; and so on all the way up. Such an approach might be attractive to legal realists, but it clearly would not have been attractive to Hart.

These issues take us into the much contested debate about judicial discretion. I shall not enter that debate much here though.[15] Instead I want to look at the question of what role legal rules can play if they are understood in this particularist way.

THE FUNCTION OF LEGAL RULES

It is often assumed that rules, especially on a positivist conception, must act as a decision procedure. Thus for example, Scott Shapiro argues that positivism is committed to the idea that rules guide judges' decisions, and Fred Schauer's book on rule-following is subtitled *A Philosophical Examination of Rule-Based Decision Making in Law and in Life*.[16] But if the kind of legal particularism that I have outlined is correct, this is a role that they can fulfil at best only partially.

In its exact sense, the idea of a decision procedure is the idea of a purely mechanical procedure for arriving at the conclusion that a given thing either does or does not have some property. In that sense decision procedures are rather hard to come by. Famously, there is not even a decision procedure for assessing the validity of an argument of the predicate calculus. But the problem with using particularist rules is much more basic than that. Even in a very loose sense of decision procedure (whatever that might be), particularist rules are clearly hopeless: one cannot start by assembling one's particularist rules and seeing how they apply to a given set of premises, since the set of potentially relevant rules is uncodifiable. We are thus far worse off than we are when trying to assess

---

[14] The term comes from Dworkin; see *Taking Rights Seriously* (London: Duckworth, 1977) pp, 32ff.

[15] It is discussed further in 'The Exception Proves the Rule'

[16] Scott Shapiro, 'On Hart's Way Out' in J. Coleman (ed.) *Hart's Postscript* (Oxford: Clarendon Press 2001); Frederick Schauer, *Playing by the Rules* (Oxford: Clarendon Press 1993).

the validity of an argument in the predicate calculus, for at least there we have the argument, and the rules of inference are known.

Partly for this reason, those who try to construct formal models in AI that mimic the reasoning employed in moral or legal domains do not employ anything like the apparatus that I have suggested: instead they use the monotonic logics that will be discussed below. My response is to concede that I am not trying to construct a decision procedure. Many people have objected that actual legal reasoning cannot be captured by an overly mechanical model; I am happy to agree with the point.

So what are the rules good for if, in many cases, they do not determine a legal decision? I suggest that they have three roles: (i) as confirmation to the agent that their initial judgment was good; (ii) as a justification to others that the decision was correct; (iii) as the basis of precedent, and hence as something that, whilst not providing anything approaching a decision procedure, will nonetheless serve as a guide.

(i) *Rules as aiding confirmation*

Don't ask for the moment how a judge arrives at a putative legal decision; just suppose that they get to it. Now they need to construct a justification for it. If the judge is bound by precedent, such a justification must work from legal rules: either existing rules, or those they create by distinguishing the case. It may be that a competent judge could construct a plausible argument for just about any conclusion, though I rather doubt that (I'll discuss it shortly). But even if that were true it doesn't show that any argument they construct is just window dressing. The argument remains is a crucial part of the judgment. And if they were unable to construct one (however unlikely that might be), their initial judgment would have to be revised.

(ii) *Rules as aiding justification*

It is scarcely controversial that legal arguments are used to justify decisions. So long as one accepts that decisions have a *ratio decidendi*, the obvious thing to identify it with is an argument, structured round a legal rule. I suggest then that the second, and perhaps most important, role of a legal rule is to structure an argument so that it provides a justification of the decision. To play this role the rules do not need to give us a decision procedure; it is enough, once the argument is constructed, that we can recognize it as good. John Dewey argued for the centrality of such a role:

> Courts not only reach decisions; they expound them, and expositions must state justifying reasons. The mental operations herein involved are somewhat different from that involved in arriving at a conclusion. The logic of exposition is different from that of search and inquiry. … [Exposition's] purpose is to set forth the grounds for the decision reached so that it will not appear as an arbitrary dictum, and so as to indicate the rule for dealing with similar cases in the future.[17]

The presence of an open-ended *That's it* clause means that it is hard to be certain that an argument is good: could there not be some other rule that one should have considered that,

---

together with certain facts would supersede the current argument? But to provide adequate justification we do not need certainty.

To whom is the justification directed? It is commonly held that courts are required to justify their actions to those they might coerce. Perhaps they do owe a justification, though as a thesis about the behaviour of courts it doesn't seem to hold generally: the decisions of French courts, for instance, are often little more than statements of the verdict. I suspect that much of the reason that courts in the common law system give such lengthy decisions is specific to that system: they are concerned to justify them to other courts, partly because they are vulnerable to appeal, and partly because they know that they are setting precedent, and so know that they need to articulate them sufficiently fully for them to serve in that role.[18]

(iii) *Rules as guides*[19]
The fact that rules cannot serve as decision procedures does not show that they have no role as a guide to decision. In fact, if the judge is bound by precedent, the rules will provide quite a tight constraint. As Grant Lamond points out in a very helpful discussion, even when a court reaffirms a rule, it thereby subtly changes it: for it affirms that the case at hand is one in which the rule does apply.[20] So a judge aiming to distinguish a case cannot pick on just any factors as distinguishing: they will need to pick on factors that have not previously been held to be consistent with the rule. That means that the application of the rule will be very tightly constrained. In most standard cases there will be no question of distinguishing, and so the rules will act as a very strong guide. Of course there will always be factors that distinguish any case from all those that have been tried before it. Many of those—the date at which the events took place, the identities of those concerned—will clearly not normally be possible grounds for distinction. And there will be many that are not as obviously bad as these, but that would none the less result in a sanction for frivolous claim were they to be raised in a court. Perhaps a sufficiently ingenious judge would always be able to find a ground to distinguish a case when they wanted to; and certainly there are cases where it is hard to see what the grounds for the distinction amounts to. But I doubt that such behaviour is common.[21]

So much then for the account I offer. I turn now to the question of whether other accounts could do better.

---

[18] For a lengthy and enlightening discussion of precedent which sees such reason giving as central within the common law tradition see Duxbury *The Nature and Authority of Precedent*, esp. Chs. 2 and 3. Duxbury goes on to say that the law should not be understood as needing to be in conformity with logic (see esp. pp. 140–9). I think that this involves an over monolithic view of logic; the important question is rather the kind of logic that the law has.

[19] This is the main place where I now feel that the argument given in 'The Exception Proves the Rule' was mistaken: I was too ready to think that rules would provide no guidance.

[20] Grant Lamond 'Do Precedents Create Rules?' *Legal Theory* 11 (2005) pp. 1–26. Lamond argues that precedent does not work by rules but rather by reasons. I think that all of the grounds he adduces for that claim can be accommodated by the account presented here.

[21] In 'The Exception Proves the Rule' I adduce some general considerations from social psychology for believing that claim. These are far from compelling. Given how hard it is to identify, it is hard to know quite how one would go about determining how much illegitimate distinguishing goes on. Alternatively, a court could simply overrule a precedent: again something that is always theoretically open, since the doctrine of stare decisis itself contains an unless-clause. Blackstone wrote that 'precedents and rules must be followed unless flatly absurd or unjust (*Commentaries* 1. p. 69)

COULD THE CLAIMS BE GENERICS?

English contains a generic construction; or, rather, a family of related generic constructions. If someone says

> Chimpanzees have ten toes; or
> The Chimpanzee has ten toes; or
> A Chimpanzee has ten toes

the claim (assuming that the same claim is made by all these sentences) is true, and it isn't made untrue by the rare cases of chimpanzees born with more, or fewer, than ten toes, or by those who have lost them in accidents. So generic sentences are sentences that admit of exceptions.

There is still much debate over how the semantics of generics works. But there is now a large body of evidence that they shouldn't be seen as universal quantification, but rather as involving some innate faculty of generalization: children understand them long before they understand universally quantified sentences.[22] Could legal rules be examples of them?

Let us start, once again, by considering moral sentences. It would be odd to use the bare plural, as in the first of the Chimpanzee sentences, and say

> Killings are wrong

But we might use the singular

> Killing is wrong.

'Killing' in that case though looks to be a verb and not a noun, since we can modify it by adding a noun:

> Killing people is wrong.

But take a case like:

> Murder is wrong

'Murder' here is clearly an abstract noun, and here we plausibly have a generic: not a count-term generic as illustrated by the Chimpanzee sentences, but a mass-term generic, like

> Water is liquid at room temperature

Note that 'murder' is already a moralized term: to count something as a murder is already to count it as at least *prima facie* wrong. It is, in other words, a thick moral (and legal) term, one containing both normative and descriptive information. And once we think of thick moral (and legal) terms we find that that examples are legion, whether critical (theft/coercion/bribery is wrong) or complimentary (loyalty/bravery/ kindness is good).

---

[22] See Sarah-Jane Leslie 'Generics: Cognition and Acquisition', *Philosophical Review* 117 (2008) pp. 1–47.

(Interestingly it is harder to find purely descriptive non-count terms that work in this way. It looks as though we may be readier to use abstract non-count terms for virtues and vices than for actions in general; but if this is so, I do not know why.)

We might well appeal to such sentences in justifying legal rules, but they are not themselves legal rules of the kind in which we are interested. Can we find legal rules that are plausibly generics? Many of what Dworkin terms 'principles' look as though they might have this form. One of his examples can be formulated as what looks like a generic, here with a bare plural:

> Wrongdoers should not benefit from their own wrongdoing.[23]

Similarly, the principle that is often said to underpin contract law is put very naturally in bare plural form:

> Agreements are to be kept[24]

As I say, these look rather like generics; I'm not sure that's right, but let's assume that they are. Still, these are not yet the kinds of rule that we find in the details of common law texts; they are too general and open-ended. As Dworkin stresses, principles can clash. An argument that contained only principles would not culminate in a conclusion for how to decide a particular case: we would need to know whether the principles applied in this case, and that requires something stronger than a principle.

So let us look further at the case of contract law. When we move beyond the general principles to something more specific we find things like this:

> A contract is only valid if it is accepted;
> but its acceptance does not count if it is not heard by the offerer (for
> instance, if it is drowned out by an aeroplane);
> but it does count if the offerer did not hear it, but it was clearly made;
> but it does not count if it was clearly made but the offerer made it known
> to the acceptor that he did not hear it.[25]

And this is just one of the necessary conditions for a contract to be valid; stating the necessary and sufficient conditions requires a whole text. These are the materials from which legal arguments are actually made. Now they do culminate in a binding conclusion, and they do so because they plausibly have the form, not of generics, but of universally quantified sentences, along the lines of:

> $\forall x$ (If x is a valid contract $\rightarrow$ (x is accepted & (the acceptance is heard by
> the offerer $\lor$ (the acceptance is clearly made & the offerer does not make
> it clear that he does not hear the acceptance))))

---

[23] Though he actually formulates it as a universally quantified claim: No man may profit from his own wrong; Dworkin, *op cit* p. 26.

[24] *Pacta sunt servanda*; here again, at least in its use in international relations, the principle is usually qualified with an unless clause: *clausula rebus sic stantibus* (things remaining as they are).

[25] Denning LJ in [1955] 2 QB 327.

Nevertheless, despite the complexity of the condition we can easily imagine circumstances in which the very considerations that prompt the complexity require further complexity: what if the offerer's statement that he did not hear the acceptance was itself drowned out by an aeroplane, and so on. So even here we need *That's it* clauses.

We can think of the development of the common law as the process of arriving at rules from a set of principles—that, I think, is a plausible account of at least one aspect of its history. If we understand principles as generics, we can therefore think of it as a move from generics to universally quantified sentences. If, in a parallel way, we think of moral principles as generics, there is a question of whether we have need of the move to moral rules at all: and therefore of whether the account sketched above for moral rules has any role to play. On the one hand, we don't have need of the rules to play the roles of public justification and of guidance that they play in the common law. On the other, if we stop just with principles then we do not moral arguments that entail their conclusions. I leave the question open.

NONMONOTONIC LOGIC

The proposal developed so far uses classical logic; indeed, it does not even involve any new connectives. An alternative, proposed in a set of important articles by John Horty, seeks to achieve a similar end by very different means.[26] A similar end, in that Horty too wants to develop a framework in which legal rules have a function although they can, at least in principle, be superseded. Different means, in that he proposes a logical framework that moves a very long way from classical logic.

Classical logic is monotonic: adding new premises to a valid argument cannot result in the argument becoming invalid. In contrast a nonmonotonic logic does allow for this possibility. So let me start by spelling out the logic that Horty wants to use, before asking what benefit is gained.

Horty suggests using a logic modeled on Reiter's default logic. Typically this is presented as an interpretation of generic sentences, but here I will consider the idea that it can serve to provide a logical interpretation for legal rules. This was developed as a logic for drawing consequences from a set of premises, given that further premises might serve to undermine those consequences. In classical first order logic understood proof-theoretically, we start with a set of sentences that are built either just from the non-logical vocabulary, or the non-logical vocabulary together with the logical vocabulary: the connectives and quantifiers. The consequences of a set of sentences are simply the sentences that follow from that set given the rules of inference, rules that operate on the original sentences in virtue of their logical form. Thus if the original set contains P and (P → Q), and *modus ponens* is one of the rules of inference, then Q will follow from the

---

[26] 'Reasons as Default's *Philosopher's Imprint* 7 (2007); 'Reasons and Precedent' ms. For discussion of the nonmonotonic logic that lies behind his discussion, see his 'Nonmonotonic Logic', *The Blackwell Guide to Philosophical Logic,* L. Goble (ed.), (Oxford: Blackwell, 2001), pp. 336–361.

original set. We can think of the *conclusion set* as being the set of sentences that is obtained by collecting all the logical consequences so understood.[27]

But classical logic is monotonic. In moving to a nonmonotonic logic, Reiter takes the premise set to consist not just of the set of sentences as classically conceived, but in addition a set of *default rules*. Although these are often referred to as rules of inference, they are not akin to classical rules of inference like *modus ponens*. Rather they are more like additional premises. We can think of them as something like a three-place conditional, roughly of the form: If A, then B, provided that C is compatible with the conclusion set. Symbolize this as

$$(A: C \ / \ B).[28]$$

We can see now why the approach is nonmonotonic. If our initial set contained just A and (A: C / B), then we would be entitled to add B to the conclusion set. But if we were to add ~C to the initial set, this making it inconsistent with C, we should no longer be entitled to add B.

How do we construct the conclusion set given this approach? The idea, of course is that, given both A and (A: C / B) in our initial set, B should end up in the conclusion set just in case ~C is not in the conclusion set. But it turns out that constructing that set is no simple matter, since, even if ~C is absent from the initial set, it might arrive in the conclusion set as a result of other inferences. As a result there is no simple iterative procedure that will take us from the initial set to the conclusion set. Indeed, on the fixed point approach that Reiter endorses, a given initial set will sometimes enable us to arrive at different putative conclusion sets—or *extensions* as he calls them.[29] So how should we determine *the* conclusion set given different consistent extensions? One approach is to simply pick an arbitrary extension as the conclusion set. Another is to say that a sentence will go in the conclusion set just in case it is in any extension (though such a set may not be consistent). A third is to say that the conclusion set contains only those sentences that appear in every extension (though such a set may be empty).[30]

Deciding which of these approaches is right is beyond the scope of our enquiry here. Let us instead focus on how the default rule approach, with its new nonmonotonic logic, gives us any advantages over the *unless*-clause approach that I have been advocating. In considering this question, Horty argues that the default logic has two things in its favor.[31]

---

[27] More standardly in classical logic we would simply call a set closed under logical entailment a *theory*. But I will try to keep the discussion parallel to Reiter's.

[28] In Horty's discussion the default rules are discussed in ways that make them seem more like classical inference rules—the idea being that given A, one is committed to inferring B provided that C is compatible with the conclusion set. But they are not logical rules; they are not supposed to apply to a sentence just in virtue of that sentence's logical form. Their application crucially depends upon the interpretation of the non-logical vocabulary. For some discussion of the issues here, and of whether some discussions of nonmonotonic logic muddle the idea of the frame and the interpretation function see Robert Stalnaker, 'What is a nonmonotonic consequence relation?', *Fundamenta Informaticae*, vol. 21, (1994) pp. 7-21. In general Stalnaker argues that it might be helpful to recast nonmonotonic logic, understood in terms of the consequence relation, as a theory of nonmonotonic operators—i.e. operators that do not obey the rule of inference (If p |= q then Op |= Oq)—within a monotonic logic.

[29] For summary of the fixed point approach see Horty, 'Nonmonotonic Logic'.

[30] In recent work Horty has reduced the indeterminacy here by introducing an ordering on default rules. But since he only requires it to be partial, indeterminacy will remain.

[31] Horty 'Nonmonotonic Logic 'p. 7

Here he has in mind as the alternative a theory that simply lists the possible *unless*-conditions as conditions on the antecedent:

If A, and if C_1 & C_2 & C_3 … then B

where C_n is the negation of the nth *unless*-condition. Call this the *list account*. It is rather different to the quantified approach involved in the *That's it* proposal that I have advocated. But it will help fix the issues if we first consider the simpler list account; since even here I am skeptical that the default rule approach brings much real advantage.

The first problem that Horty identifies with the list account is that 'the list of circumstances that might interfere … is open-ended. No conceivable list of possible interfering circumstances could be complete'[32] That is a legitimate worry, and it is the reason that I tried to capture the defeating conditions using quantifiers rather than a list. But does the default rule account escape the worry? Defeating conditions only get into the default approach once they are incorporated into the default theory. One obvious way of incorporating them is to make the default rule look like this:

(A: C_1 & C_2 & C_3 … / B)

If the worry is that no finite list account could list all of the possible defeating conditions, then it is equally true that no finite default rule could capture them either. In both cases the theory that we have will only accommodate some of the interfering conditions. We can always extend it by adding more—more clauses in the antecedent for the list account, more clauses in the justification for the default account—but the two are on a par.

The default approach is more flexible than this though. Rather than trying to incorporate the interfering conditions directly into the default rule, we could write that just as

(A: B/ B)

meaning that we can add B to the conclusion set provided that we do not also have ~B in it; and then the defeating conditions can be represented by further premises like

If ~C_1 then ~B
If ~C_2 then ~B …

Or by further default rules like

(~C_1: D/~B)
(~C_2: D/~B) …

But this doesn't change the central point. You still can't get something for nothing. If some factor is going to play a role as a defeating condition in the default theory, it can

32 *Ibid.*

only do so if it is explicitly introduced at some point. And that makes it just as vulnerable to unlistable defeat conditions as the list approach.

Let us turn then to the second factor. Horty writes:

> The second problem is more subtle, and would arise even if we did have relatively exhaustive list of qualifications. The point of placing preconditions in the antecedent of a [principle] is that we must verify that the preconditions are satisfied before concluding that the [principle applies]… But it seems less reasonable to suppose that we must actually have to verify that all of the various weird circumstances that might interfere with this [principle] do not occur … It would be better to be able simply to assume that weird circumstances like these do not occur unless there is information to the contrary.[33]

This was an important consideration in the development of default logic—it is the initial motivation in Reiter original article for instance—so we should spend some time on it. Part of the reason for providing default logics was to provide a logic that enabled one to move forwards—to draw conclusions, or to perform actions—in the absence of information. Default logics don't give us a decision procedure for determining membership of the conclusion set.[34] But the idea was that, to take the standard example, on learning that Tweety is a bird, we should be able to defeasibly conclude that Tweety can fly. We do not want to have to hold off on that conclusion while we make sure that Tweety is not a penguin, or an emu, or an ostrich or has had his wings clipped, or was born deformed, or whatever. But clearly if we had built these conditions into the antecedent of the conditional

> If Tweety is a bird, & Tweety is not a penguin & Tweety is not an emu…
> then Tweety can fly

we would need to establish each of the conjuncts of the antecedent before we could detach the consequent.

In contrast, in the absence of defeating conditions, the default approach will enable us to move forward. We do not need to establish that Tweety is not a penguin, not an emu and so on. Provided that we do not have evidence that he is, the instantiation of the default rule

> (x is a bird: x can fly / x can fly)

will enable us to (defeasibly) conclude, from the information that Tweety is a bird, that Tweety can fly.

---

[33] *Ibid.*

[34] Indeed, Reiter's default logic is even worse off then first order predicate calculus, in that it is not even semi-decidable, i.e. not only is there no procedure for demonstrating that a sentence isn't in the set, there is also no procedure for showing that a sentence is. See Reiter pp. 104ff.

Call this the *no stalling* feature of the default account. Sometimes no stalling will indeed be a feature we want. Typically whether we do so will depend upon the number of exceptions that there are to the rule, the benefits of moving on and the costs of the occasional mistake. In making rough and ready predictions about a creature's flying capacity a no stalling feature might be very useful. Is it something we want in moral and legal reasoning though?

This is a difficult question. In moral reasoning I think that it generally isn't. We put high store on getting the conclusion right. By and large, if there are known defeating conditions, we want people to ensure that they are not met before the move to a moral judgment. In the terms of default approach, we ask that they check that none of the known defeating conditions obtain; and that involves adding to the initial evidence set information on whether they are or are not. So it seems to me that the no stalling feature is not much of an advantage in moral reasoning. In fact, it can look to be something of a liability.

In legal reasoning though the situation is rather different. Courts and jurisdictions differ, but in the adversarial Anglo-American system, it isn't normally the role of the judge and jury to seek out new information; decision are made on the basis of the (admissible) evidence presented. To that extent then the default account looks to give a good basis for legal reasoning. For the court needs to not stall: it needs to make a decision on the basis of the evidence it has.

Even here the advantages of the default account over the list account should not be overplayed. Courts have a complex set of default assumptions stemming from the idea of the burden of proof. They are allowed to assume many things unless proved otherwise, and perhaps this would allow them to complete the list. But I shall not pursue the question of whether this will work. For it strikes me that both the list approach and the default approach share a fundamental flaw.

So far we have been concerned with recognized defeating conditions are met: with clauses that are contained in the list or embedded in the default rules. When it comes to moral and legal reasoning though, we ask for more than that. We ask that agents be aware of considerations that are relevant to the judgments they make but that are not currently recognized. Admittedly when it comes to moral reasoning, this is to place the bar very high: it is only a somewhat idealized moral judge who will recognize the significance of factors that have not been recognized before. But in the case of legal reasoning the requirement is far more realistic. It is quite common for a case to come before a court that involves factors that have not previously been dealt with by the law. And in such a case, as we said before, the court will have either to reaffirm the existing rule, thereby in effect judging that the new factors are not sufficient to warrant a different verdict; or will have to distinguish.

Let us take the case where the court reaffirms the existing rule. It is important to note that here, just as much as in the case where they distinguish, courts are open to criticism. Most obviously, if the case goes to appeal, the decision may be overturned. What are we to make of such a happening? On either the list approach or the default approach, it looks as if they did nothing wrong: they applied the existing rules, and arrived at their conclusion. It might be held that applying the rules is not good enough. But if that is so, in virtue of what is it not good enough. It seems that the flaw must stem from some further theory that we have not been given.

In contrast, on the account I am offering, it is very clear where the mistake lies. In endorsing a legal argument, the court is implicitly endorsing the *That's it* condition. They are committed to the claim that there are no further relevant facts and rules. And in the circumstances imagined, that was not true.

A parallel issue arises when the court does distinguish, and where on appeal the move is rejected. Here again, we want to know what it was that the court did wrong. And again the default approach has nothing to say on the matter, whereas the approach that I have offered does: the court held that the *That's it* clause in the argument using the original rule was false, whereas it was not.

Can we ever know whether the That's it condition is met? The issue would gain importance if we agreed with Horty's implicit requirement that we must *verify* the premises in an argument. The *That's it* condition is equivalent to a universally quantified sentence, and Karl Popper built a hugely influential theory on the idea that they can never be verified. We need not follow him in that to think that we are in no position to verify that the *That's it* condition is met. But what I think we should conclude is that verification is far too high a standard for the premises in legal argument. As with much else in law, we might want them to be beyond reasonable doubt. However, we want to concede that judges might make mistakes; and this strikes me as placing the fallibility in just the right place.

In responding to Horty I have made no use of the familiar idea that classical logic is well behaved and well understood, whereas nonmonotonic logic is not. I do think that that is true; and so if two theories did the job equally well, I think that we would still have grounds to prefer a theory that used classical logic over one that did not. But since I am sceptical that the default theory does the job equally well, I have had no need of the argument.

CONCLUSION

In relation to its rivals, the *that's it* approach stands up surprisingly well. It does require us to posit implicit claims, but they are far from outlandish. Perhaps it has little role to play in moral argument: that will depend how well a generic account can be developed. But in law, where we do have need of tight, precedence justifying arguments that nonetheless admit of exceptions, I know of no better approach.